

BESZÉDKUTATÁS 2015

**Szerkesztette:
Gósy Mária**

MTA Nyelvtudományi Intézet
Kempelen Farkas Beszédkutató Laboratórium
Budapest, 2015

A szerkesztőbizottság elnöke: Gósy Mária (MTA Nyelvtudományi Intézet)
A szerkesztőbizottság tagjai: Bunta Ferenc (University of Houston)
Horváth Viktória (MTA Nyelvtudományi Intézet)
Mády Katalin (MTA Nyelvtudományi Intézet)
Markó Alexandra (Eötvös Loránd Tudományegyetem)
Olaszy Gábor (Budapesti Műszaki és Gazdaságtudományi Egyetem)
Siptár Péter (Eötvös Loránd Tudományegyetem)
Vago, Robert (Queens College és City University of New York)

A korpusz szerkesztésében részt vettek: Bóna Judit
Gyarmathy Dorottya
Markó Alexandra
Vakula Tímea

Az egyes tanulmányokat szakterületi kompetenciával rendelkező szakemberek lektorálták.

Technikai szerkesztés: Neuberger Tilda és Markó Alexandra

A kötet kiadását az MTA Könyv- és Folyóiratkiadó Bizottsága támogatta.

A folyóiratot az MTMT indexeli és a REAL archiválja.

DOI: 10.15775/Beszkut.2015.23.

ISSN 1218-8727

© MTA Nyelvtudományi Intézet

Felelős kiadó: Kenesei István igazgató

Készült: Ook-Press Kft., Veszprém

TARTALOM

Neuberger Tilda: Zöngétlen zárhangok időszerkezete a fonológiai hosszúság függvényében.....	5
Honbolygó Ferenc – Kolozsvári Orsolya: A hangsúly észlelésének akusztikai meghatározói.....	21
Kohári Anna – Markó Alexandra: A megnyilatkozás végének jelzése felolvasásban: temporális szerkezet és zöngeminőség.....	35
Gósy Mária – Krepsz Valéria: Magánhangzók temporális jellemzői az idő múlásának függvényében	53
Bárkányi Zsuzsanna – G. Kiss Zoltán: The laryngeal properties of Slovak three-consonant clusters.....	66
Gyarmathy Dorottya: Az <i>úgyhogy</i> funkcióbővülése a spontán beszédben	92
Bóna Judit: Nonverbális hangjelenségek fiatalok és idősek spontán beszédében	106
Laczkó Krisztina – Tátrai Szilárd: A metapragmatikai tudatosság jelzései számítógép közvetítette társalgási narratívákban	120
Beke András – Horváth Viktória: Kisiskolások alaphangmagasságának variabilitása	133
Tar Éva: Szó eleji zárhangok zöngékezdesi ideje: beszédproduktions adatok az óvodás korosztályra vonatkozólag	148
Auszmann Anita: A magánhangzók akusztikai szerkezete 9 és 11 éves iskolások spontán beszédében.....	164
Alkalmazások, eszközök.....	176
Csapó Tamás Gábor – Csopor Dávid: Ultrahangos nyelvkontúrkövetés automatikusan: A mély neuronhálókra alapuló AutoTrace eljárás vizsgálata	176
Hunyadi László – Szekrényes István – Sziklai István: Vizuális percepció és nyelvi feldolgozás	186
Gyarmathy Dorottya – Neuberger Tilda: Egy hiánypótló adatbázis: a Tini BEA	209
Olaszy Gábor – Abari Kálmán: Az artikulációs mozgások akusztikai vetületének adatbázisa magyar beszédre.....	222

Németh Géza – Zainkó Csaba – Bartalis Mátyás – Olaszy Gábor: Többnyelvű vasúti hangos utastájékoztatók kórusz alapú TTS- módszerrel.....	233
Klessa Katarzyna – Markó Alexandra: Annotation Pro: Új eszköz a beszéd nyelvi és paranyelvi sajátosságainak címkézésére és elemzésére	242
„Nyelvbtlás”-kórusz.....	248
Téves szó.....	249
Grammatikai hiba.....	250
Kontamináció	251
Téves kezdés	251
„Nyelvem hegyén van” jelenség	252
Perszeveráció	253
Anticipáció	253
Metatézis	254
Egyszerű nyelvbtlás	254
Több típusba sorolható jelenségek	255
A magyar és a nemzetközi tudományos élet eseményei	257
Gombocz Zoltán (1877–1935) • <i>Tóth Andrea</i>	257
Egy fizikus, aki megalapozta a hazai akusztikát, és kedvenc kutatási területe a beszédhangok vizsgálata volt. Idén lenne 100 éves Tarnóczy Tamás (1915–2007) • <i>Vicsi Klára</i>	259
10 th International Seminar on Speech Production • <i>Ausmann Anita</i>	261
Workshop on Late Stages in Speech and Communication Development • <i>Ausmann Anita</i>	262
A <i>Beszéd – Kutatás – Alkalmazás</i> című sorozat • <i>Nagy-Varga Zsolt</i>	263
A kötet szerzői	267

ZÖNGÉTTLEN ZÁRHANGOK IDŐSZERKEZETE A FONOLÓGIAI HOSSZÚSÁG FÜGGVÉNYÉBEN

Neuberger Tilda

Bevezetés

A nyelvtudomány két különböző ága, a fonológia és a fonetika a nyelvi rendszer egyazon szintjén lévő egységek kutatásával foglalkozik – egyik fő céljuk a nyelv fonémáinak/ beszédhangjainak egzakt leírása –, ábrázolásmódjuk azonban eltérő. A fonológiai reprezentáció elsődlegesen kvalitatív, kognitív, és önmegfigyelésen is alapulhat, ezzel szemben a fonetikai jellemzően kvantitatív, fizikailag hozzáférhető és mérhető, valamint kevésbé introspektív (Pierrehumbert 1990). A két tudományág azonban számos közösnek tekinthető kérdéskörrel foglalkozik. A határterületen megjelenő kérdések egyike abból fakad, hogy a fonológiai megkülönböztető jegyek, amelyek egy adott nyelv magánhangzó- és mássalhangzórendszerét meghatározzák, a beszédhangok szintjén is realizálódnak az artikulációban, az akusztikumban és a percepcióban. A magyar mássalhangzók artikulációs konfigurációja például négy paraméterrel adható meg, amelyek alapján az adott mássalhangzó egyértelműen definiálható: a képzés módja, a képzés helye, a hangszalagok működése (zöngéesség) és a nyelvi időtartam (Gósy 2004). A fonetikai rendszerezés szerinti jegyek fonológiai is relevánsnak tekinthetők. A mássalhangzókészlet elemeinek ábrázolására a klasszikus generatív fonológiai keretben a megkülönböztető jegyes elemzést (a jegymátrixos formát), míg az autoszegmentális fonológiai keretben a geometrikus jelölésmódot szokták alkalmazni (Siptár 1995). A fonológiai ábrázolásban például a [szon] jegy alapján különböztethetők meg az obstruensek és a szonoránsok (amely a képzés módjáról ad információkat), vagy a [zöng] jegy alapján a zöngés és a zöngétlen mássalhangzók (amely a hangszalagok működéséről ad információkat) (Vago 1980).

A hangidőtartam szerepe háromféleképpen jelenhet meg a nyelvi rendszerben: a) fiziológiailag determinált, fonológiai irreleváns; b) nyelvi feltételektől függően automatikusan létrejövő, nyelvi funkciók ellátására nem képes, redundáns; c) nyelvi kód által determinált, disztinktív funkcióval rendelkező (Kassai 1979). A beszédhangok időtartama a világ nyelveinek többségében a magánhangzórendszeren belül jut nagyobb nyelvi szerephez. Több olyan nyelv van, amelyben a magánhangzók hosszúság szerint oppozícióban állnak, a mássalhangzók viszont nem; ennek fordítottja azonban ritkább. (Ezen megállapításunkat az UPSID adatbázis 451 nyelve alapján tettük, vö. Maddieson 1984.) Bizonyos nyelvekben a mássalhangzó-időtartam morfémaán belül nem

kontrasztív. Ilyen például az angol, amelyben gemináták ('ikerhangok', kettőzött mássalhangzók) csak morféma határán jöhetnek létre (például *top pick* : *topic*); ellentétben a japánnal, amelyben a morféma belüli rövid és hosszú mássalhangzók minimálpárokat alkotnak (például *oto* 'hang' : *otto* 'férj') (Hayes 2001). Delattre (1971) négy nyelv geminátáinak akusztikai és artikulációs korrelátumait vizsgálta angol, német, spanyol és francia anyanyelvű beszélők rögzített anyagában. A gemináták mind a négy nyelvben előfordulhatnak morfémahatáron, például angol *will lend* : *will end*; német *Stiehl Loden* : *Stiehl Oden*; spanyol *el lecho* : *el hecho*; francia *Il l'aime* : *Il aime* (a példák az említett tanulmányból származnak). Az eredmények szerint a rövid mássalhangzót és a geminátát tartalmazó szavak elkülönítésében nemcsak az érintett mássalhangzók időtartambeli, hanem azok intenzitáskülönbségei is szerepet játszanak.

A magyarban a nyelvi kvantitást releváns (kontrasztív) fonológiai jegynek szokás tekinteni, amelynek disztinktív szerepe erőteljesen érvényesül a magyar magánhangzórendszerben, a mássalhangzórendszerben azonban csak korlátozottan jelenik meg. A mássalhangzó-hosszúság mint fonológiai jegy leggyakrabban olyan minimális párokat különböztet meg, amelyek egyik vagy mindkét tagja a) levezetett (morfémahatáron vagy hasonulással keletkezett) szóban, például *vasal* : *vassal*; b) hangutánzó, hangulatfestő vagy indulatszóban, például *csat* : *csatt*; c) jövevényszóban vagy mozaikszóban, például *kasza* : *kassza*; d) tulajdonnévben fordul elő, például *Bereg* : *berreg* (Siptár 2001). A mögöttes, tehát a szótárban is ilyen alakban szereplő (például *ép* : *épp*), nem levezetett szembenállások alacsony terheltsége periferikus jelenséggé teszi a fonológiai kvantitást (Obendorfer 1975; Siptár 2012). Ebből kifolyólag a magyar mássalhangzórendszerben a kvantitásoppozíciót számos tanulmány szükségtelennek tartja, a rendszertakarékosság végett a hosszú mássalhangzókat két azonos mássalhangzó kapcsolatának, geminátának tekintik (pl. Deme 1971; Siptár 2001). A hosszúság disztinktív jegyét nem a fonéma tulajdonságaként kezeli, hanem a morfémaéként. A hosszú mássalhangzók morfofonológiai voltát támasztja alá az is, hogy polifonematikusak (azaz két fonéma értékűek), ugyanis szótaghatáron a két elemük külön-külön szótagba kerül (Deme 1971). Az ikeredés szabályát a kis előfordulású, mögöttes hosszú mássalhangzót tartalmazó szavakra is alkalmazva a mássalhangzórendszer terjedelme felére csökkenthető (Siptár 2001). Ettől függetlenül az így keletkező gemináták a fonetikai ábrázolásban hosszú felszíni mássalhangzóként jelennek meg.

A levezetett alakok egy része fonológiai koartikuláció és/vagy lenizációs folyamatok eredményeképpen keletkezik. A természetes fonológia elméleti keretében a lenizációs (simító) folyamatok közé olyan folyamatok tartoznak, amelyek az artikuláció kényelmességével függenek össze. Ezek a beszélő és a hallgató szempontjait figyelembe véve (vagyis funkcionális osztályozás szerint) a beszélő szempontjait részesítik előnyben – a szegmentumok artiku-

lációjának leegyszerűsítésével (Kiefer 2001). A magyarban ilyen lenizációs folyamat az (időértékeket is befolyásoló) hasonulás (például *adta* [ɒtɒ]) vagy a mássalhangzó-rövidülés/degemináció (például *add meg* [ɒdmɛg]) (Nádasdy 1989). Az egy képzési jegyet érintő hasonulás során az egyik hang a másikhoz hasonlóvá válhat a hangszalagműködés tekintetében (zöngésségi hasonulás) (Gósy 2004). A folyamat eredményeképpen a két szomszédos fonéma akár teljesen azonos beszédhanggá válhat (például *fogkefe* [fok:efe]). Létrejöhhetnek teljesen azonos beszédhangok alaktani kötöttségű hasonulás folyamán is, mint ahogyan az a *-val/-vel*, *-vá/-vé* toldalékok vagy az *az*, *ez* mutató névmás ragozott formái esetében tapasztalható (például *fokkal* [fok:ɔl]). Általában egyetlen hosszú mássalhangzó jön létre a felszínen akkor is, ha morféma- (pl. *hattól*) vagy szóhatáron (pl. *hat tojás*) két azonos mássalhangzó kerül egymás mellé (az ikeredés szabálya szerint) (Siptár 2001). Ez azonban felveti a gemináta, az álgemináta és a hosszú mássalhangzó fogalmak közötti különbség kérdését.

A nemzetközi szakirodalomban a *gemináta* terminust alkalmazzák általánosan a hosszú mássalhangzókra, a hazai szakirodalomban azonban a múlt század közepétől kezdve találhatunk olyan munkákat, amelyek megkülönböztetik a *gemináta* és a *hosszú mássalhangzó* terminusokat, artikulációs szempontból, illetőleg a fonetikai realizáció tekintetében eltérő jelenségeket értenek alattuk (pl. Hegedűs 1956; Fónagy–Fónagy 1969). Ezek szerint a két fogalom közötti különbség abban rejlik, hogy míg a hosszú mássalhangzókra az „egységes artikuláció, nyújtott ejtés, emelkedő-ereszkedő szájüregi nyomás és izomfeszültség” jellemző, addig a geminátákra a kettős artikuláció, esetükben a regisztrátumon az izomfeszültség és a szájüregi nyomás két csúcsot jelez”, vagyis mintha egymás után két hang képzését követhetnénk nyomon (Kassai 1979: 16). Ilyen utóbbi fonetikai realizációkat találtak magyar, angol, német, spanyol, francia és észt nyelven is (Fónagy–Fónagy 1969; Delattre 1971; Lehiste et al. 1973). Az úgynevezett álgemináták esetében (vagyis amikor két azonos rövid mássalhangzó kerül egymás mellé morféma- vagy szóhatáron) nem ritka a kettős artikuláció, illetve szünet beillesztése a két hang közé (például az *un-* előjárós angol szavak esetében) (Földi 1989; Oh–Redford 2012). Mindazonáltal mind a valódi, mind az álgemináták hosszabb időtartamértékeket mutattak, mint a rövid mássalhangzók (pl. Oh–Redford 2012).

Fonetikai szempontból (a jelen tanulmány is elsődlegesen ebből a szempontból vizsgálja a jelenséget) a gemináták hosszú mássalhangzóknak tekinthetők. A rövid-hosszú mássalhangzók fizikai időtartamértékei nagy átfedéseket mutatnak, és arányuk nyelvenként igen változatos lehet (Ham 2012). Ladefoged és Maddieson (1996) vizsgálata szerint az egyes nyelvekben a hosszú mássalhangzók hossza átlagosan másfélszerese-háromszorosa rövid párjukénak. A japán zár- és réshangok akusztikai elemzésekor azt találták, hogy az időtartam-különbség a rövid és a hosszú explozívák között nagyobb (/t/-/tt/ között 180 ms), mint a rövid és a hosszú frikatívák között (/s/-/ss/ kö-

zött 134 ms) (Hayes 2001). Az olasz nyelvben ennek ellenkezőjét bizonyították; a hosszú explozívák időtartama rövid párjuk 201%-a volt, míg ez az arány a frikatívák esetében 173%-os (Giovanardi–Di Benedetto 1998).

A magyar nyelv mássalhangzóit tekintve a rövid-hosszú arány különféle értékeket mutat a különböző kutatásokban. Olasz (2006) vizsgálatában a hosszú mássalhangzók átlagos hossza a rövidekének 165%-a. Egy másik elemzésében (Olasz 2007) az intervokális helyzetű rövid mássalhangzók átlagosan 92 ms időtartamúak, a hosszúak pedig 162 ms-osak voltak (vagyis arányuk 176%-osnak bizonyult). Az alveoláris és a posztalveoláris zöngétlen résmássalhangzók nyelvileg rövid és hosszú előfordulásai között szignifikáns különbséget mutatott ki Beke–Gyarmathy (2010); akiknél a rövid [s]-realizációk átlagos időtartama 91 ms, a hosszúaké pedig 120 ms volt (vagyis arányuk 132%-osnak mutatkozott), a rövid [ʃ]-realizációk átlagosan 89 ms-osak, a hosszúak pedig 130 ms-osak voltak (vagyis arányuk 146% volt). Grácz (2012) vizsgálatában a hosszú zöngés explozívák szó belseji intervokális helyzetben 20–50%-kal realizálódtak hosszabban rövid párjuknál, míg a zöngétlenek közül a /p/ esetében a hosszú fonéma megvalósulásai lényegében nem tértek el a rövidekétől, a /t/ és a /k/ esetében pedig 20–30%-os volt az időtartam-különbség. Szó belseji helyzetben vizsgálták a /t/ explozíva, a /ʃ/ réshang és a /tʃ/ affrikáta nyelvileg rövid és hosszú realizációit, és azt találták, hogy mindhárom konzonáns szignifikánsan hosszabban valósul meg geminátaként, mint egyes (vagyis rövid) mássalhangzóként (Pycha 2007). Ugyanabban a tanulmányban azt is igazolták, hogy a hosszú affrikáták esetében mind a zár-, mind a réselem időtartamában tapasztalható időtartamnyúlás, a záralkotás szakaszában azonban ez jelentősebb mértékű. A felpattanó zárhangok esetében a fonológiai kvantitás meghatározó paramétere a zárszakasz hosszában keresendő. Sem a pozitív VOT, sem a felpattanás ideje nem támogatja akusztikailag a rövid-hosszú kontrasztot. Ezt több nyelven is alátámasztották: olasz (Esposito–Di Benedetto 1996), török és bengáli (Lahiri–Hankamer 1988; Hankamer et al. 1989), levantei arab, magyar, madurai (Indonézia, Jáva szigete) (Ham 2012).

A jelen tanulmány a fonológiai rövidnek tekinthető zöngétlen explozívák időszerkezetét elemzi, és hasonlítja össze a felszínen megjelenő hosszú mássalhangzók időértékeivel. A fő kutatási kérdés az, hogy a három elemzett zárhang /p, t, k/ esetében a beszédhangok teljes időtartamában és belső időviszonyaiban milyen különbségek mutathatók ki a fonológiai hosszúság szerint. Elemzésünk kiterjed a hosszú mássalhangzók különböző típusaira is, abban az értelemben, hogy azok fonológiai természetüket tekintve milyen kategóriába sorolhatók: mögöttes, valódi levezetett vagy álgemináták. Hipotézisünk szerint a különböző geminátatípusok felszíni megvalósítása fonetikailag azonos, legalábbis az időtartamuk alapján nem egyértelműen elkülöníthetők egymástól.

Kísérleti személyek, anyag, módszer

A kísérlethez a BEA adatbázisból (Gósy et al. 2012) választottuk ki hét beszélő hanganyagát. Mindegyik adatközlő egynyelvű, magyar anyanyelvű, budapesti férfi. Életkoruk 20 és 29 év közötti (átlagéletkor: 24,1 év). A spontánbeszéd-felvételben munkájukról, hobbijukról, valamint egy, a felvételvezető által megadott témáról (pl. ünnepek, iskolai emlékek, magántulajdon védelme) beszéltek kötetlenül.

A vizsgálat anyagát 855 zöngétlen felpattanó zárhang képezte, amelyek az 1. táblázatban foglaltak szerint oszlottak meg a képzési hely és az időtartam alapján. A képzési hely szerint összesen 154 bilabiális, 376 alveoláris és 325 veláris explozívát elemeztünk, közülük 589 volt rövid, 266 pedig hosszú mássalhangzó. Egy beszélőtől átlagosan 122 beszédhangot címkéztünk fel. A hosszú /t/ és /k/ esetében minden beszélőtől legkevesebb tíz előfordulást adatoltunk. A /p/ előfordulása spontán beszédben ritkának bizonyult, azon adatközlők eredményeit vetettük statisztikai elemzés alá, akiknél legalább öt nyelvileg hosszú mássalhangzó-előfordulást találtunk (egy adatközlőnél egy, míg egy másiknál egy sem jelent meg a hosszú, zöngétlen, bilabiális explozívából).

1. táblázat: A vizsgált zárhangok megoszlása a képzési hely és a fonológiai hosszúság szerint

A vizsgált mássalhangzó	Fonológiai hosszúság		Összes (db)
	Rövid (db)	Hosszú (db)	
/p/	123	31	154
/t/	240	136	376
/k/	226	99	325

Mivel morfémán belül a hosszú mássalhangzók csak intervokális helyzetben vagy szó végén fordulhatnak elő, és a szó végi hosszú mássalhangzók egy esetleges (jobbról) szegélyező mássalhangzó hatására megrövidülhetnek (degeminálódhatnak) (l. Grácz–Siptár 2013), az összes előfordulást úgy választottuk ki, hogy a hosszú mássalhangzót mindkét oldalról magánhangzó szegélyezze (V_V vagy V_#V).

Egyetlen hosszú beszédhangnak tekintettünk minden, a felszínen megjelenő hosszú mássalhangzót, legyen akár mögöttes gemináta (pl. *éppen*, *kettő*, *csökken*), valódi levezetett gemináta (pl. *kalappal*, *attól*, *dolgokkal*) vagy álgemináta (pl. *diákként*, *művészettörténet*, *másik kettő*). A /tt/ múlt idő jelét Vago (1980) alapján mögöttes alakként értelmezhetjük (pl. *látott*), amely egy másik mássalhangzót követően degeminálódik (pl. *várt*). Siptár–Törkenczy (2000) szerint azonban levezetett valódi geminátáról van szó, amely akkor jön létre (az úgynevezett *t*-terjedési szabály által), ha magánhangzót követ a múlt idő jele. Elemzésünkben az utóbbi álláspontot érvényesítettük. A valódi levezetett gemináták között adatoltunk még zöngésségi hasonulás kimenete-

ként létrejövő hosszú mássalhangzókat is, például *tudtam*, *légkört*. A 266 hosszú mássalhangzó közül 57 darab mögöttes (morfémán belüli), 197 darab levezetett valódi és 12 darab álgeminátát elemeztünk.

A vizsgált fonetikai paraméterek között a mássalhangzók teljes időtartama, a zárszakasz időtartama, a zárszakasznak a teljes időtartamhoz viszonyított aránya (zárarány), a zöngékezdési idő (VOT) és a felpattanások száma szerepeltek. Abban az esetben, amikor felpattanás nem volt adatolható, csak a beszédhang teljes időtartamát vizsgáltuk.

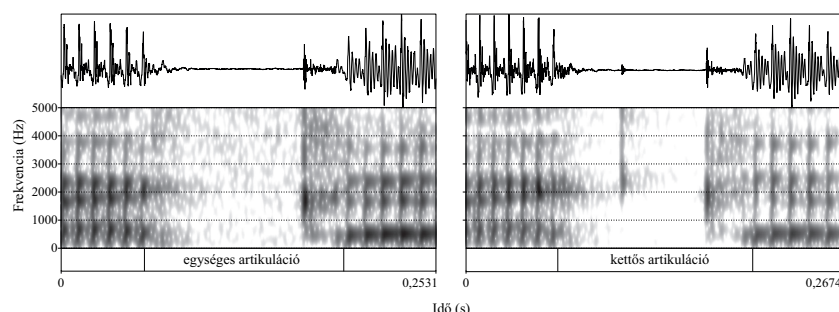
A címkézést és a mérést a Praat 5.3 verziójának szoftverrel végeztük (Boersma–Weenink 2013). Az annotálás manuálisan történt a spektrogram és az oscillogram mintázata alapján, auditív ellenőrzéssel. Magánhangzó-alapú szegmentálást alkalmaztunk, vagyis a vizsgált zárhang határait a megelőző, illetve a követő magánhangzó második formánsa alapján jelöltük. A többszöri felpattanások esetében a zárszakasz végét, valamint a VOT-értékeket az első felpattanástól számítottuk (még ha nem is minden esetben ez volt a legintenzívebb felpattanás).

Az adatokat statisztikai próbáknak vetettük alá, amelyekhez az SPSS 13.0 verziójának szoftvert használtuk. A rövid és a hosszú mássalhangzók időértékei közötti különbségek szignifikáns voltát beszélőkön belül vizsgáltuk a nem paraméteres, összetartozó mintás Wilcoxon-próbával (a Monte-Carlo sztochasztikus szimulációs módszerrel kiegészítve). A geminátatípusok normalizált időértékei közötti különbségeket pedig a nem paraméteres Kruskal–Wallis-próbával, valamint a Mann–Whitney *U*-próbával teszteltük. A statisztikai elemzéseket 95%-os konfidenciaintervallumon végeztük el.

Eredmények

A bevezetésben említett hosszú mássalhangzó vs. gemináta megkülönböztetés szerint az elemzés anyagát képező hosszú konsonánsok túlnyomó többsége (99,25%-a) hosszú mássalhangzóként, egységes artikulációjú, nyújtott hangként realizálódott. Mindössze két előfordulást (0,75%) adatoltunk az úgynevezett kettős artikulációra, vagyis amikor a regisztrátumon a gemináta két egymást követő hangként valósult meg. Mindkét előfordulás zöngésségi hasonlással keletkezett, levezetett gemináta esetében jelent meg (egyik a *tömegközlekedés*, másik a *kiadták* szóban). Az egységes és a kettős artikuláció akusztikai vetületét hasonlítja össze az 1. ábra a *tömegközlekedés* szó kétféle realizációjában. Mindkét előfordulást egyazon beszélő produkciójában adatoltuk. Míg az egységes artikuláció esetében egy felpattanás látható a regisztrátumon, és a zárszakasz relatíve hosszú, addig a kettős artikulációra jellemző volt a két külön felpattanás.

A fonológiai hosszúság fonetikai realizációjának kérdéskörében a továbbiakban a vizsgált zárhangok teljes időtartamát, zárszakaszuk időtartamát, valamint a VOT értékeit vetjük össze a rövid és a hosszú mássalhangzók esetében.



1. ábra

Egységes és kettős artikuláció a *tömegközlekedés* szó /gk/ kapcsolatában

Teljes időtartam

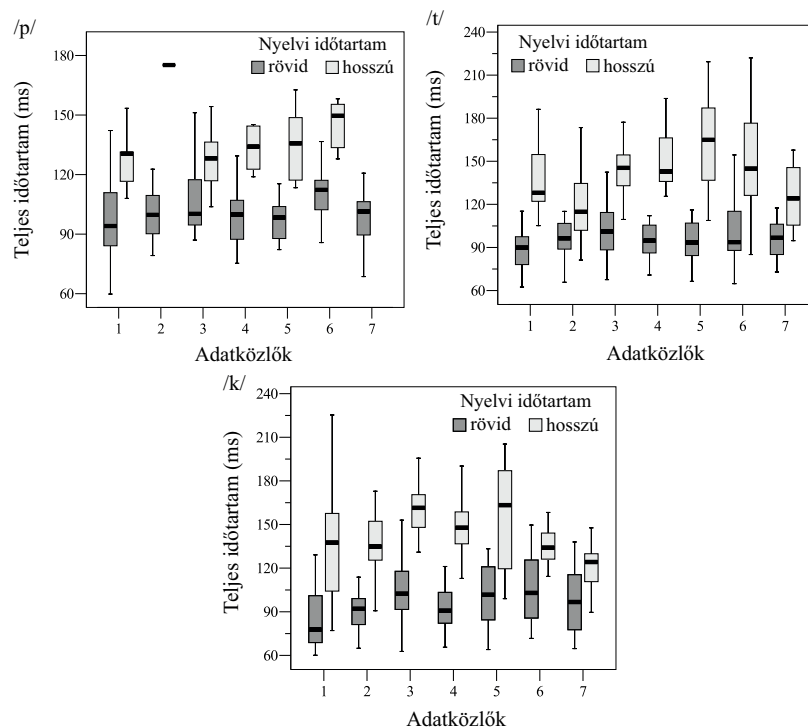
A három vizsgált explozív teljes időtartamának átlaga és szórása a hét beszélőnél a következőképpen alakult. A rövid mássalhangzók esetében: a [p] 101 (± 16) ms, a [t] 95 (± 19) ms, a [k] 98 (± 22) ms, míg a hosszú mássalhangzók esetében: a [p:] 137 (± 20) ms, a [t:] 143 (± 31) ms, a [k:] 143 (± 29) ms időtartamú volt. Az összetartozó mintás, nemparaméteres Wilcoxon-próba szignifikáns különbséget mutatott a nyelvileg rövid és hosszú mássalhangzók teljes időtartamértékei között; a /p/ esetében ($Z = -2,201$; $p = 0,028$; Monte-Carlo-szimulációval: $p = 0,033$), a /t/ esetében ($Z = -7,646$; $p = 0,018$; Monte-Carlo-szimulációval: $p = 0,015$) és a /k/ esetében $Z = -10,052$; $p = 0,018$; Monte-Carlo-szimulációval: $p = 0,015$) is. Elvárásainknak megfelelően mind a hét adatközlőnél a hosszú mássalhangzók időértékei bizonyultak hosszabbnak a rövidekéhez képest (2. ábra).

Beszélőnként megállapítottuk a rövid-hosszú mássalhangzók teljes időtartamának átlagértékeit, majd kiszámoltuk arányukat, vagyis azt, hogy a hosszú mássalhangzók időtartama hányszorosa a rövidekének. A bilabiális zöngétlen explozív esetében a nyelvileg hosszú realizációk időtartama a rövid párjukénak 121–176%-a (az összes beszélő átlagértékét tekintve 140%-os), ez az arány az alveoláris explozív esetében a beszélők között 129–179% (átlagosan 150%), a veláris explozívánál pedig 126–167% (átlagosan 149%). Ezek az eredmények azt mutatják, hogy a hosszú mássalhangzók időértéke a rövid párok időtartamának körülbelül másfélszerese.

A zárszakasz időtartama

A zárszakasz időtartamának átlaga és szórása a hét beszélőnél a következő értékeket mutatta. A rövid mássalhangzók esetében: [p] 79 (± 14) ms, [t] 71 (± 18) ms, [k] 63 (± 18) ms, míg a hosszú mássalhangzók esetében: [p:] 115 (± 20) ms, [t:] 122 (± 31) ms, [k:] 106 (± 27) ms időtartamú volt. Szignifikáns különbséget találtunk a nyelvileg rövid és hosszú mássalhangzók zárszakaszának időtartama között (Wilcoxon-próba) a /p/ esetében ($Z = -2,201$; $p = 0,028$; Monte-Carlo-szimulációval is: $p = 0,035$), a /t/ esetében ($Z = -2,366$;

$p = 0,018$; Monte-Carlo-szimulációval is: $p = 0,017$) és a /k/ esetében $Z = -2,366$; $p = 0,018$; Monte-Carlo-szimulációval is: $p = 0,017$) is.

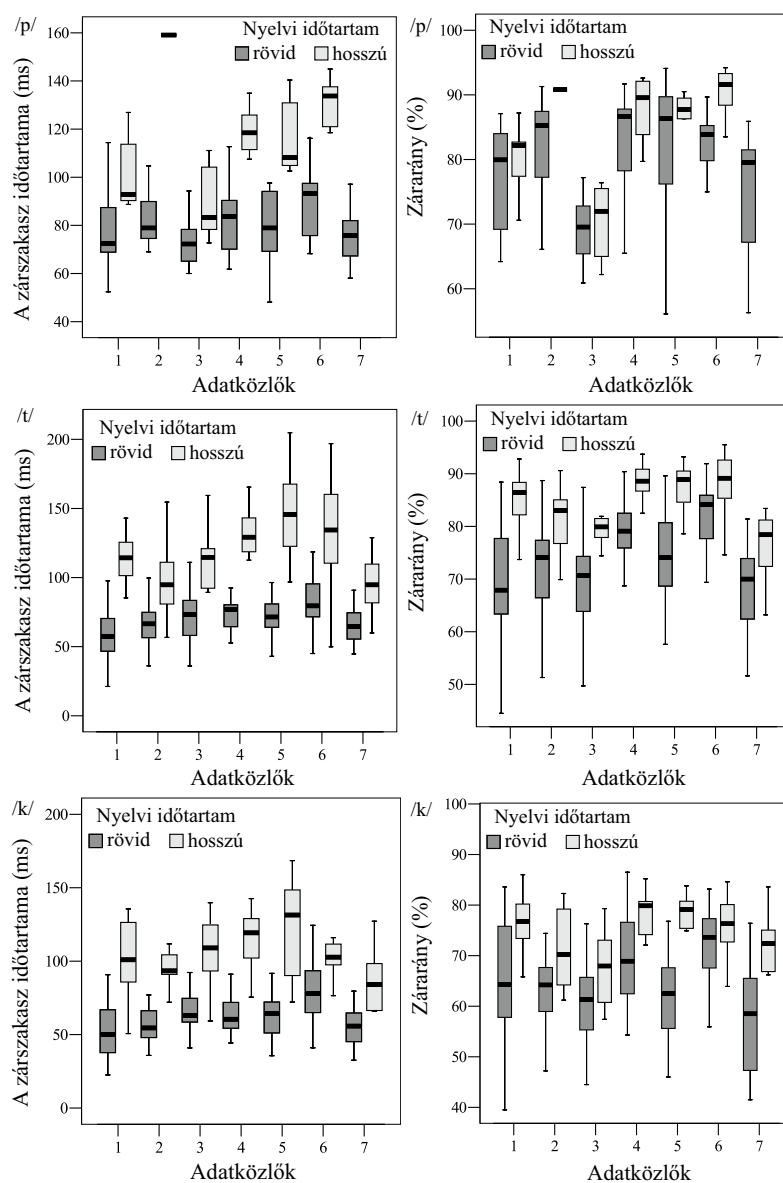


2. ábra

A nyelvileg rövid és hosszú explozívák teljes időtartama (ms) beszélőnként

A hosszú mássalhangzók zárszakaszának időtartamában megnyúlást tapasztaltunk a rövid párjukéhoz képest (3. ábra), ez a nyúlás a [p] hangnál a beszélők között 122–197%-os (átlagosan 149%-os) volt, a [t] hangnál 145–196%-os (átlagosan 171%-os), a [k]-nál pedig 135–194%-os (átlagosan 170%-os). A hosszú beszédhangok zárszakasza a képzési helytől függetlenül vizsgálva (vagyis mindhárom zárhangot együttevén) átlagosan 164%-ban nyúlt meg a rövid explozívák zárszakaszának időtartamához képest.

Megvizsgáltuk, hogy a zárszakasz időtartama hogyan aránylik a beszédhang teljes időtartamához (3. ábra). A zárarány a fonológiai rövid beszédhangoknál átlagosan 78,6%-os volt a [p], 73,2%-os a [t], és 64,1%-os a [k] esetében, míg a hosszú zárhangoknál átlagosan 84,0% a [p:], 83,6%-os a [t:], és 73,7%-os a [k:] esetében.



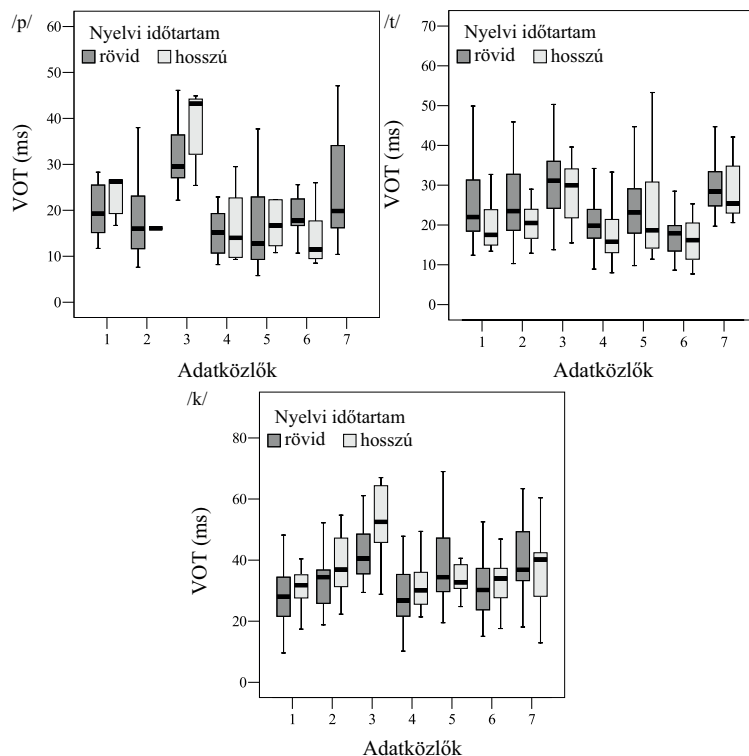
3. ábra

A nyelvi rövid és hosszú explozívák zárszakaszának időtartama (ms) és a zárarány (%) beszélőnként

A rövid és a hosszú zárhangok zárárányában tapasztalt különbség mind a három hang esetében szignifikáns; a /p/ ($Z = -2,201$; $p = 0,028$; Monte-Carlo-szimulációval is: $p = 0,035$), a /t/ ($Z = -2,366$; $p = 0,018$; Monte-Carlo-szimulációval is: $p = 0,015$) és a /k/ ($Z = -2,366$; $p = 0,018$; Monte-Carlo-szimulációval is: $p = 0,015$) esetében is. Ezek az adatok azt mutatják, hogy nemcsak a hosszú mássalhangzók teljes időtartama nőtt meg – magával vonva a zárszakasz időtartamának megnyúlását, hanem arányaiban is változás történt a zárhangok belső időszerkezetében.

A zöngekezdési idő

A VOT-átlagok és a szórások a hét beszélőnél hasonló értékeket mutattak a rövid és a hosszú hangok esetében (4. ábra). A rövid mássalhangzóknál: [p] 22 (± 11) ms, [t] 25 (± 10) ms, [k] 35 (± 13) ms, míg a hosszú mássalhangzók esetében: [p:] 22 (± 13) ms, [t:] 22 (± 10) ms, [k:] 37 (± 13) ms volt a VOT.



4. ábra

A nyelvileg rövid és hosszú explozívák zöngekezdési ideje (ms) beszélőnként

A statisztikai elemzés (Wilcoxon-próba) csak a /t/ esetében jelzett szignifikáns különbséget a rövid és a hosszú hangok zöngékezdési idejében ($Z = -2,366$; $p = 0,018$; Monte-Carlo-szimulációval: $p = 0,016$). A másik két zárhang rövid-hosszú párijainak esetében nem mutatható ki különbség a VOT-értékekben. Érdekes eredmény, hogy a hosszú [t:] hangok VOT-értékei minden adatközlőnél rövidebb időtartammal realizálódtak, mint a rövid [t] VOT-értékei; az arány a beszélők között 77–98%-os (átlagosan 90%-os). A másik két mássalhangzó esetében jóllehet nem volt statisztikailag igazolható időtartam-különbség, a hosszú hangok zöngékezdési ideje bizonyult hosszabbnak, a /p/-nél a rövidekének 104%-a, a /k/-nél pedig 107%-a volt.

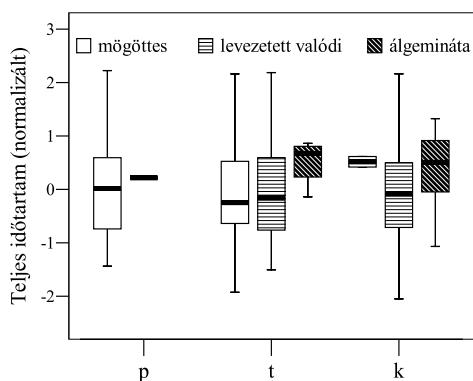
A különböző geminátatípusok időviszonyai

Annak érdekében, hogy a különböző geminátatípusok időértékeit össze tudjuk hasonlítani, az adatokat beszélőnként normalizáltuk az adott beszélő összes hosszú mássalhangzójának időtartamátlagához és -szórásához. Ezzel kívántuk kiküszöbölni az egyéni ejtési sajátosságokból, például a különböző artikulációs tempóértékekből fakadó beszélők közötti eltéréseket. A normalizálást a három beszédhangra külön-külön végeztük el, hiszen a képzési hely is befolyásolja az időértékeket. Összevetettük az egyes geminátatípusok normalizált teljes időtartamát (5. ábra), zárszakaszának időtartamát (6. ábra), valamint zöngékezdési idejét (7. ábra). A [p:] beszédhangokat a kis adatmennyiség miatt kizártuk a statisztikai elemzésből (álgeminátát egyáltalán nem adatoltunk belőle), de az ábrákon szemléltetésképpen megjelenítettük az időértékeit.

A mássalhangzók teljes időtartamát tekintve a /t/ esetében a legrövidebben a mögöttes gemináták (átlag: 143 ms), a leghosszabban pedig az álgemináták (átlag: 162 ms) realizálódtak; a /k/ esetében pedig a legrövidebb időtartamúak a levezetett valódi gemináták (átlag: 141 ms), leghosszabbak pedig a mögöttes gemináták (átlag: 175 ms) voltak (5. ábra). A statisztikai elemzés alapján nem találtunk szignifikáns különbséget a különböző típusú gemináták teljes időtartama között egyik mássalhangzó esetében sem (Kruskal–Wallis-teszt, $p > 0,05$).

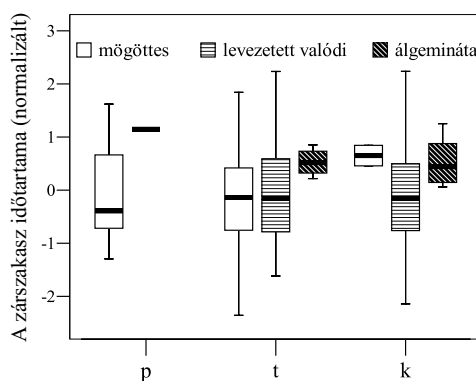
A gemináták zárszakaszának időtartama az egyes típusok közül a /t/-nél a legrövidebb a mögöttes geminátáknál (átlag: 118 ms) volt, míg a leghosszabb zárszakaszok az álgeminátákra (átlag: 141 ms) voltak jellemzőek; a /k/-nél – a teljes időtartamhoz hasonlóan – a levezetett valódi gemináták zárszakasza volt a legrövidebb (átlag: 104 ms), és a mögötteseké a leghosszabb (átlag: 133 ms) (6. ábra). A statisztikai elemzés ebben a vonatkozásban nem mutatott ki szignifikáns eltérést a /t/ esetében (Kruskal–Wallis-teszt, $p > 0,05$), a /k/ esetében azonban szignifikáns különbséget találtunk ($\chi^2 = 6,159$; $p = 0,046$; Monte-Carlo-szimulációval is: $p = 0,030$). Az utóbbi mássalhangzónál az egyes csoportok közötti összehasonlítást a Mann–Whitney U -teszttel végeztük, ami szignifikáns eltérést jelzett a levezetett valódi és az álgemináták között ($Z = -2,131$; $p = 0,033$). Mindent összevetve, mind az alveoláris, mind

a veláris zöngétlen zárhang esetében az adatok az álgemináták hosszabb teljes és zárszakasz-időtartamára utalnak a levezetett valódi geminátákkal szemben. A mögöttes gemináták akusztikai korrelátumait vizsgálva azt találtuk, hogy míg a [t:] hangot illetően a levezetett valódi geminátákra hasonlít jobban, addig a [k:] hangnál az időértékek közelebb állnak az álgemináták értékeihez.



5. ábra

A különböző geminátatípusok teljes időtartama

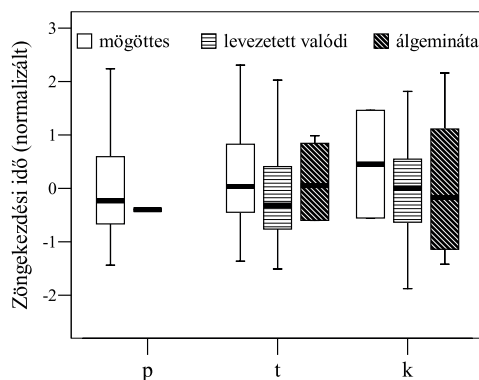


6. ábra

A különböző geminátatípusok zárszakaszának időtartama

Végül elemeztük a VOT-értékeket is annak érdekében, hogy lehetséges különbségeket adatoljunk a három geminátatípus között. A zöngekezdési idő a három típus közül az álgemináták esetében bizonyult a legrövidebbnek (átlag: /t/ 21 ms és /k/ 34 ms), és a mögötteseknél a leghosszabbnak (átlag: /t/ 24 ms és /k/ 43 ms) (7. ábra). Matematikailag igazolható különbséget azonban nem

találtunk közöttük (Kruskal–Wallis-teszt, $p > 0,05$), ami abból is adódhat, hogy a VOT értéke a követő magánhangzó minőségétől is függ (vö. pl. Gósy 2000), ezt azonban a jelen kutatásban nem elemeztük.



7. ábra

A különböző geminátatípusok zöngékezdési ideje

Következtetések

Tanulmányunkban a magyar zöngétlen explozívák időszerkezetét vizsgáltuk a fonológiai hosszúság függvényében. Arra a kérdésre kerestük a választ, hogy a fonológiai hosszúság és a gemináció típusa hogyan, milyen mértékben jelentkezik a fonetikai időviszonyok megváltozásában.

Az egységes artikulációval szemben a gemináták kettős artikulációja eleynesző számú volt vizsgálatunk anyagában. Kijelenthetjük, hogy az egységes, nyújtott artikulációt tekinthetjük a hosszú mássalhangzók tipikus ejtésének. Meg kell jegyeznünk azonban, hogy az ettől eltérő, kisszámú előfordulás a hasonulással keletkezett geminátákat érintette. Ezen (tipikustól eltérő) realizációk háttérében két lehetséges magyarázat állhat. Egyfelől megemlíthetjük a hiperartikuláció jelenségét (Lindblom 1990), amely bizonyos mássalhangzókapcsolatokban a zöngésségi hasonulás szabályos megvalósulását is akadályozhatja, esetenként a két érintett hang közötti szünet beiktatásával (vö. Gósy 1999; Markó et al. 2010). Másfelől beszédtervezési, illetve kivitelezési nehézség (a fonológiai és/vagy az artikulációs tervezés szintjén bekövetkező bizonytalanság, illetve összehangolatlanság) is gátolhatta a hosszú mássalhangzó egységes artikulációs megvalósítását (vö. Frisch–Wright 2002).

A három zöngétlen zárhang elemzése során azt találtuk, hogy a fonológiai oppozíció jelölésére szolgáló időtartam-növekedés legkevésbé a [p] explozíva esetében jelent meg, a [t] és a [k] beszédhangra nagyobb mértékben volt jellemző. Ez a megállapítás Grácsi (2012) eredményeivel is összhangban áll, amelyek szerint a bilabiális rövid-hosszú zöngétlen explozívák időtartama

nem tér el jelentősen (rövid-hosszú átlagértékek: 110 ms, illetve 116 ms), a veláris zárhangok esetében nagyobb az eltérés (rövid-hosszú átlagértékek: 102 ms és 129 ms), a legnagyobb különbség pedig a rövid-hosszú alveoláris zöngétlen explozív esetében mutatkozik (rövid-hosszú átlagértékek: 98 és 139 ms). Eredményeink szerint a hosszú mássalhangzók teljes időtartama mintegy másfélszerese a homorgán rövid zárhangokénak. A megnyúlás a felpattanó zárhangok első időszaki részai közül a zárszakaszt érintette. A zárszakaszok időtartamának meghosszabbodása a teljes időtartamhoz hasonlóan a [t] és a [k] zárhangok esetében nagyobb arányú volt (átlagosan 171%-os, illetve 170%-os), mint a [p] esetében. A zöngékezdési idő értékei túlnyomórészt változatlanok bizonyultak annak függvényében, hogy az adott beszédhang nyelvileg rövid vagy hosszú. Mindebből arra következtethetünk, hogy a fonológiai hosszúság jelölésére a zárszakasz időtartam-növekedése szolgál (képzési helytől függően különböző mértékben), ellenben a zöngékezdési idő többé-kevésbé változatlan marad, így a fonológiai kvantitás kifejezésében ez utóbbi paraméter nem tölt be jelentős szerepet.

A különböző geminátatípusok között az objektív időértékek alapján nem találtunk jelentős különbségeket, vagyis hipotézisünk, miszerint felszíni megvalósításuk fonetikailag azonos, beigazolódott. Tendenciaszerűen megjelent ugyan, hogy az álgemináták teljes időtartamban, illetve a zárszakaszuk időtartamában hosszabban realizálódtak, mint a mögöttes vagy a levezetett valódi gemináták, a zöngékezdési idejük pedig rövidebb volt, mint a másik két típusé, a statisztikai elemzés azonban nem jelzett szignifikáns különbségeket. Minde mellett az egyes típusok kevés előfordulása miatt csak óvatos következtetéseket tehetünk, a megbízhatóbb eredmények érdekében nagyobb adatmennyiségre van szükség. Továbbá meg kell említenünk, hogy a jelen eredmények az adatközlők választott csoportjából kifolyólag a fiatal felnőtt férfi beszélőkre vonatkoznak; általánosabb képet akkor kaphatunk, ha a jelenségkört a női beszélőkre, más életkorokra is kiterjesztjük. Az esetlegesen megjelenő nemek közötti eltéréseket következő kutatásainkban tervezzük vizsgálni.

A fonetika és a fonológia egyik közös problémaköre arra vezethető vissza, hogy egyfelől a fizikai időtartam kvantitássá (nyelvileg hasznosított hangjellemzővé) válhat, másfelől a kvantitás a hangképzést kísérő fizikai hangtulajdonsággá alakulhat (Kassai 1979: 19). Ebből adódóan felmerül a kérdés, hogy a fizikai időtartam miként vetíthető le tartamra, vagyis a fonológiai kvantitás fokaira. A kvantitás az idő síkján elemezhető jelenségeken túl milyen más paraméterekben érvényesül (pl. formánsszerkezet, alaphangfrekvencia változása)? Hogyan lehet összhangba hozni/egymásra vonatkoztatni az esz-közfonetikai elemzés objektív értékeinek sokaságát az egynemű fonológiai kategóriákkal? Az ilyen és hasonló, határterületen lévő kérdések megválaszolására további vizsgálatok szükségesek, amelyek a két tudományterület eredményeit egymást kiegészítve értelmezik.

Irodalom

- Beke András – Gyarmathy Dorottya 2010. Zöngétlen résmássalhangzók akusztikai szerkezete. *Beszédkutatás* 2010. 57–75.
- Boersma, Paul – Weenink, David 2013. Praat: doing phonetics by computer [Computer program]. 5.3. verzió. <http://www.praat.org/> (A letöltés ideje: 2013. október 10.)
- Delattre, Pierre 1971. Consonant gemination in four languages: An acoustic, perceptual, and radiographic study. Part I. *International Review of Applied Linguistics in Language Teaching* 9/1. 31–52.
- Deme László 1971. A magyar fonémák rendszeréhez és rendszerezéséhez. *Néprajz és Nyelvtudomány* 15–16. 89–99.
- Espósito, Anna – Di Benedetto, Maria Gabriella 1996. Acoustical and perceptual study of gemination in Italian stops. *The Journal of the Acoustical Society of America* 106/4. 2051–2062.
- Fónagy Iván – Fónagy Éva 1969. Szájüregi nyomásmérések. In Pais Dezső – Benkő Loránd (szerk.): *Dolgozatok a hangtan köréből*. Nyelvtudományi Értekezések 67. Akadémiai Kiadó, Budapest. 17–44.
- Földi Éva 1989. Magánhangzó-nazalizáció, hosszú mássalhangzó vagy gemináta? Hozzászólás Bolla Kálmán: A magyar hangtan időszerű problémái c. előadásához. *Egyetemi Fonetikai Füzetek* 2. 98–120.
- Frisch, Stefan A. – Wright, Richard 2002. The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics* 30/2. 139–162.
- Giovanardi, Maurizio – Di Benedetto, Maria-Gabriella 1998. Acoustic analysis of singleton and geminate fricatives in Italian. *The European Student Journal of Language and Speech* 1. 1–15.
- Gósy, Mária 1999. Predictability of voicing assimilation in speech production. In Ohala, John J. – Hasegawa, Yoko – Ohala, Manjari – Granville, Daniel – Bailey, Ashlee C. (eds.): *Proceedings of the 14th International Congress of Phonetic Sciences* (San Francisco). Vol. 3. University of California, Berkeley. 1905–1908.
- Gósy Mária 2000. A /p, t, k/ mássalhangzók zöngékezdesi ideje. *Magyar Nyelvőr* 124/2. 195–203.
- Gósy Mária 2004. *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest.
- Gósy Mária – Gyarmathy Dorottya – Horváth Viktória – Grácsi Tekla Etelka – Beke András – Neuberger Tilda – Nikléczy Péter 2012. BEA: Beszélt nyelvi adatbázis. In Gósy Mária (szerk.): *Beszéd, adatbázis, kutatások*. Akadémiai Kiadó, Budapest. 9–24.
- Grácsi Tekla Etelka 2012. *Zöreghangok akusztikai fonetikai vizsgálata a zöngésségi oppozíció függvényében*. Doktori disszertáció. ELTE, Budapest.
- Grácsi Tekla Etelka – Siptár Péter 2013. Degemináció? *Beszédkutatás* 2013. 7–25.
- Hankamer, Jorge – Lahiri, Aditi – Koreman, Jacques 1989. Perception of consonant length: Voiceless stops in Turkish and Bengali. *Journal of Phonetics* 17/4. 283–298.
- Ham, William H. 2012. *Phonetic and phonological aspects of geminate timing*. Routledge, New York.
- Hayes, Rachel L. 2001. The perception of novel phoneme contrasts in a second language: A developmental study of native speakers of English learning Japanese singleton and geminate consonant contrasts. In Hayes, Rachel – Lewis, W. D. – O'Bryan, E. L. – Zamuner, T. S. (eds.): *Coyote Working Papers 12. Language in Cognitive Science*. University of Arizona. 28–41.

- Hegedűs Lajos 1956. A geminátákról. *Magyar Nyelv* 52. 415–428.
- Kassai Ilona 1979. *Időtartam és kvantitás a magyar nyelvben*. Nyelvtudományi Értekezések 102. Akadémiai Kiadó, Budapest.
- Kiefer Ferenc 2001. A fonológia ma. In Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 2. Fonológia*. Akadémiai Kiadó, Budapest. 25–41.
- Ladefoged, Peter – Maddieson, Ian 1996. *The sounds of the world's languages*. Blackwell, Oxford.
- Lahiri, Aditi – Hankamer, Jorge 1988. The timing of geminate consonants. *Journal of Phonetics* 16/3. 327–338.
- Lehiste, Ilse – Morton, Katherine – Tatham, Mark A. 1973. An instrumental study of consonant gemination. *Journal of Phonetics* 1. 131–148.
- Lindblom, Björn 1990. Explaining phonetic variation: A sketch of the H&H theory. In Hardcastle, William J. – Marchal, Alain (eds.): *Speech production and speech modeling*. Kluwer Academic Publishers, Dordrecht. 403–440.
- Maddieson, Ian 1984. *Patterns of sounds*. Cambridge University Press, Cambridge.
- Markó Alexandra – Grácsi Tekla Etelka – Bóna Judit 2010. The realisation of voicing assimilation rules in Hungarian spontaneous and read speech: Case studies. *Acta Linguistica Hungarica* 57/2–3. 210–238.
- Nádasdy, Ádám 1989. The exact domain of consonant degemination in Hungarian. *Hungarian Papers in Phonetics* 21. 104–107.
- Obendorfer, Rudolf 1975. The ambiguous status of Hungarian long consonants. *Lingua* 36. 325–336.
- Oh, Grace E. – Redford, Melissa A. 2012. The production and phonetic representation of fake geminates in English. *Journal of Phonetics* 40/1. 82–91.
- Olaszy Gábor 2006. *Hangidőtartamok és időszerkezeti elemek a magyar beszédben*. Nyelvtudományi Értekezések 155. Akadémiai Kiadó, Budapest.
- Olaszy Gábor 2007. *Mássalhangzó-kapcsolódások a magyar beszédben*. Segédkönyvek a nyelvészet tanulmányozásához 72. Tinta Könyvkiadó, Budapest.
- Pierrehumbert, Janet 1990. Phonological and phonetic representation. *Journal of Phonetics* 18/3. 375–394.
- Pycha, Anne 2007. Phonetic vs. phonological lengthening in affricates. In: *Proceedings of the 16th International Congress of Phonetic Sciences*. University of Saarland, Saarbrücken. 1757–1760.
- Siptár Péter 1995. *A magyar mássalhangzók fonológiája*. MTA Nyelvtudományi Intézet, Budapest.
- Siptár Péter 2001. A mássalhangzók. In Kiefer Ferenc (szerk.) *Strukturális magyar nyelvtan 2. Fonológia*. Akadémiai Kiadó, Budapest. 183–272.
- Siptár Péter 2012. Tényleg van-e a magyarban degemináció? In Markó Alexandra (szerk.): *Beszédtudomány. Az anyanyelv-elsajátítástól a zöngékezdési időig*. ELTE BTK – MTA Nyelvtudományi Intézet, Budapest. 19–34.
- Siptár, Péter – Törkenczy, Miklós 2000. *The phonology of Hungarian*. Clarendon Press – Oxford University Press, Oxford.
- Vago, Robert 1980. *The sound patterns of Hungarian*. Georgetown University Press, Washington.

A HANGSÚLY ÉSZLELÉSÉNEK AKUSZTIKAI MEGHATÁROZÓI

Honbolygó Ferenc – Kolozsvári Orsolya

Bevezetés

A beszéd észlelése és megértése során észlelőrendszerünk két, alapvetően eltérő típusú akusztikai információt dolgoz fel: a szegmentális, vagyis a beszédhangok, hangkapcsolatok, hangsorok jellemzőire vonatkozó, és a szupraszegmentális (prozódiai), vagyis a több beszédhangon átívelő, hangsorok egészét moduláló információt (Gósy 2004). A jelen tanulmányban a szupraszegmentális jellemzők közül a hangsúly, ezen belül is a szóhangsúly percepcióját vizsgáltuk, és arra a kérdésre kerestük a választ, hogy milyen akusztikai információk befolyásolják a hangsúly észlelését.

A hangsúly bizonyos szótagok relatív kiemelése a szavakon belül vagy bizonyos szavak kiemelése a mondatokon belül. Funkcióját tekintve kulminatív vagy demarkatív szerepet játszik, azaz a beszédfolyam bizonyos részeit kiemeli vagy elválasztja (l. Kager 2007). A hangsúly a pszicholingvisztikai szakirodalomban elsősorban a lexikai hozzáféréssel kapcsolatos elméletekben jelenik meg. Például a Metrikai Szegmentációs Elmélet (Cutler–Norris 1988) szerint a lexikai hozzáférés, azaz a nyelvi inputból történő jelentés kinyerése során a hangsúlyos szótagoknak az lehet a szerepe, hogy jelöljék a szavak kezdetét, ahonnan a lexikai hozzáférés elindulhat.

A hangsúllyal kapcsolatos percepciós elméletek kiindulópontja a hangsúly akusztikai-fonetikai jellemzőinek mérése volt. Az első elképzelések szerint a hangsúlyos szótagot nagyobb produkciós erőfeszítéssel képezzük, ezért a hangsúly legfontosabb feltételezett korrelátuma az intenzitás (Sweet 1906; Bloomfield 1933). Fónagy Iván szerint: „[a fonetikusok abból] a feltevésből indultak ki, hogy az erőteljesebben ejtett zöngé képzésekor a hangszalagok erősebb rezgésbe jönnek, és ennek megfelelően a levegőrészecskék rezgési amplitúdója is fokozódik. Ha ez így van, úgy a hangsúlyos szótagokat nagyobb amplitúdók jelzik majd a kimogrammmokon is” (Fónagy 1958: 4). Ezt az elképzelést Fónagy saját vizsgálatai nem támasztották alá, nem sikerült bizonyítania, hogy a hangsúlyos szótag minden esetben nagyobb amplitúdóértékkel járna együtt, mint a hangsúlytalan (Fónagy 1958).

A hangsúly akusztikai jellemzőit vizsgáló további tanulmányok ugyanakkor pozitívabb eredményekre vezettek. Sluijter és van Heuven (1996) holland szavakban vizsgálta a hangsúly akusztikai korrelátumait. A vizsgálat korpuszát tíz holland beszélő által felmondott, hangsúly minimális párok képezték.

Ezek olyan szó párok, amelyek ugyanazon fonémákból állnak, és csak hangsúlymintázatukban térnek el egymástól, például *'permit* (engedély) – *per'mit* (engedélyezni). A szerzők a szótagok időtartamát, átlagos intenzitását, valamint a magánhangzók formánsszerkezetét és spektrális egyensúlyát mérték. Ez utóbbi akusztikai jellemző az egyes frekvenciatartományokban mérhető energia nagyságát tükrözi. Az eredmények szerint a hangsúly minimális párokat legmegbízhatóbban elkülönítő akusztikai jellemző az időtartam volt, ezt követte a spektrális egyensúly a magasabb frekvenciatartományokban, majd az intenzitás és a formánsszerkezet.

Campbell és Beckman (1997) a hangsúly akusztikai korrelátumait különböző fókusz típusokkal rendelkező angol nyelvű mondatok esetében vizsgálta. Kutatásukban négy személy olvasott fel különböző intonációs szerkezettel rendelkező mondatokat, és a szerzők az átlagos alaphangfrekvenciát, az időtartamot és a spektrális egyensúlyt vizsgálták. Sluijter és van Heuven (1996) eredményeihez hasonlóan a magasabb frekvenciatartományok spektrális egyensúlyában találtak eltérést a hangsúlyos és hangsúlytalan szótagok között, de csak azon szavak esetében, amelyek fókuszpozícióban álltak. Campbell és Beckman (1997) ezt a különbséget az előző tanulmány eredményeihez képest azzal magyarázza, hogy a két vizsgált nyelv (angol és holland) eltér a hangsúly észlelhetőségének tekintetében. Az angolhoz képest a hollandban kevesebb olyan szó van, amelyben a hangsúlytalan szótagok magánhangzói minőségileg is változnak (azaz redukálódnak, l. svá). Az angol tehát nemcsak szuprasegmentális, hanem szegmentális jellemzőkkel is kifejezheti a hangsúlyt, és a szerzők szerint ez állhat annak hátterében, hogy a hangsúlyos és hangsúlytalan szótagok közötti akusztikai különbség csak a fókuszpozícióban álló szavak esetében volt kimutatható.

Plag, Kunter és Schramm (2011) több szótagú szavak esetében vizsgálta az elsődleges és másodlagos hangsúly lehetséges akusztikai korrelátumait, nevezetesen az időtartamot, az átlagos intenzitást, az átlagos alaphangfrekvenciát, a hanglejtést és a spektrális egyensúlyt az angolban. Eredményeik szerint az f_0 , az intenzitás és a spektrális egyensúly mutatóiban volt eltérés a hangsúly két megjelenési formája között.

A hangsúly akusztikai jellemzőinek mérésével kapcsolatos vizsgálatok tehát azt találták, hogy létezik ugyan néhány olyan jellemző, amely a hangsúlyos és hangsúlytalan szótagot elkülöníti egymástól, de ezek eltérnek az egyes nyelvek között, és az egyéb tényezők (fókusz, elsődleges-másodlagos hangsúly) jelentősen befolyásolják ezeket. A hangsúly percepciójával kapcsolatos vizsgálatok éppen ezért főként szintetizált beszédet használtak annak megállapításához, hogy mely akusztikai jellemzők járulnak hozzá az észlelésben a hangsúlyos és hangsúlytalan szótagok elkülönítéséhez. Fry (1958) hangsúly minimális párokat hozott létre az alaphangfrekvencia, az intenzitás és az időtartam manipulálása révén, és a résztvevők feladata annak eldöntése volt, hogy a bemutatott öt, hangsúlymintázat alapján igeként és főnévként is ér-

telmezhető szó melyik változatát hallották. A szintetizált szavak esetében öt lépésben változtatták az egyes akusztikai jellemzőket. Az eredmények szerint a hangsúlyítéletekre a legnagyobb hatással az időtartam változása volt, az intenzitás ugyanakkor önmagában nem volt elég a hangsúlyélmény megváltoztatásához. Az f_0 változása kapcsán a szerzők egy „minden vagy semmi” hatást találtak: ez a változás nagyságtól függetlenül megváltoztatta a hangsúlyítéletet.

Turk és Sawusch (1996) a *mama* szóban változtatták lépésenként a szótagok időtartamát és intenzitását úgy, hogy vagy az első, vagy a második szótagon változtatták az értékeket (ebben a kísérletben az f_0 -érték konstans volt a szó teljes hosszában). Fry (1958) eredményéhez hasonlóan a szerzők azt találták, hogy a hangsúlyítéleteket elsősorban az időtartam határozza meg, és a hallgatók az intenzitás változására csak kismértékben támaszkodnak. A vizsgálat eredményei azt is kimutatták, hogy az időtartam-információt a hallgatók könnyebben kinyerték az akusztikai bemenetből, mint az intenzitást.

Sluijter, van Heuven és Pacilly (1997) szintén az időtartam és intenzitás változtatásának hatását vizsgálták a hangsúlyítéletekre. Korábbi vizsgálataikra (Sluijter–van Heuven 1996) alapozva a spektrális egyensúlyt is bevették a manipulált akusztikai jellemzők körébe. Kísérletükben a *nana* ismétlődő vagy reiteratív álszót használták. Úgy vélték, hogy az ilyen típusú ismétlődő szótagokat tartalmazó reiteratív szavak azért szolgálhatnak ideális nyelvi anyagnak a percepció vizsgálatokhoz, mert lehetővé teszik, hogy kizárólag a különböző hangsúlyhoz kapcsolódó akusztikai jellemzőket módosítsuk anélkül, hogy a szegmentális jellemzőkből származó f_0 -, időtartam- és intenzitásváltozásokat figyelembe vennénk. Saját vizsgálatunkban hasonló okok miatt használtuk mi is ugyanezt a reiteratív álszót. Sluijter és munkatársai (1997) egy férfi beszélő által felmondott szövegből választottak ki egyetlen hangsúlytalan *na* szótagot, majd ezt a szótagot megkettőzték, összefűzték, és ezen szótagok szerkesztésével és újrasyntetizálásával hozták létre a kísérleti nyelvi anyagot. Az akusztikai jellemzőket hét lépésben változtatták, melyek során az egyes jellemzők nagysága az első szótagon fokozatosan csökkent, a másodikon pedig fokozatosan nőtt, és ennek megfelelően a hangsúly az első szótagról a másodikra tevődött át. A résztvevők feladata annak eldöntése volt, hogy a *nana* álszó mely szótagját hallották hangsúlyosnak egy kényszerválasztásos feladatban. Az eredmények szerint az összes hangsúlyítélet 57%-a az első szótagra esett, azaz a résztvevők a valódinál nagyobb gyakorisággal vélték úgy, hogy az első szótag volt a hangsúlyos (trochaikus torzítás). Az akusztikai jellemzők tekintetében mind az időtartam, mind az intenzitás, mind a spektrális egyensúly változtatása hatással volt a hangsúlyítéletekre, de eltérő mértékben. Az időtartam változása az előző kísérletekhez hasonlóan megbízható jelzése volt a hangsúlynak, a szótagok teljes intenzitásának változása pedig szintén az előzőekhez hasonlóan gyenge jelzésként működött. Ugyanakkor a spektrális egyensúly változtatása, azaz az intenzitás specifikus

növelése a magasabb frekvenciatartományokban (500 Hz felett) fontos meghatározó jellemzőnek bizonyult. A szerzők feltételezése szerint ennek oka, hogy az intenzitás frekvenciaspecifikus változása realisztikusabb akusztikai megvalósulása a hangsúllyal együtt járó nagyobb produkciós erőfeszítésnek, mint a szótagok intenzitásának globális növekedése. Ezzel tulajdonképpen a hangsúllyal kapcsolatos legkorábbi elképzeléseket (Sweet 1906; Bloomfield 1933) próbálják rehabilitálni, és amellet érvelnek, hogy legalábbis az angol nyelvben a hangsúly megvalósulása mégiscsak valamilyen módon a szótagok nagyobb intenzitásával jár együtt, igaz, ez a korábbi feltételezésekhez képest specifikusabb módon történik.

A jelen tanulmányban hasonló kérdéseket vizsgáltunk a magyar nyelvvel kapcsolatban. Két alapvető kérdésünk volt: befolyásolja-e a hangsúlyítéleteket az, hogy a hangsúlyt milyen akusztikai jellemző hordozza, illetve megjelenik-e az első szótag preferenciája, azaz a trochaikus torzítás a hangsúlyítéletekben.

A kísérlet során a tisztán perceptuális folyamatokat kívántuk vizsgálni, és ennek érdekében a nyelvi környezetből kiemelt, jelentés nélküli álszavak észlelését vizsgáltuk. Ezt oly módon tettük, hogy változtattuk a hangsúly pozícióját a hangsúlyhoz kapcsolódó akusztikai jellemzők mesterséges módosítása révén, és arra kértük a résztvevőket, hogy hozzanak döntést a hangsúlyos szótag pozíciójával kapcsolatban. Noha ez távol áll a hétköznapi beszédészlelési folyamattól, mégis értékes információval szolgálhat azon alapvető észlelési mechanizmusokról, amelyek a hangsúly észlelését alátámasztják.

A vizsgálatban három különböző, a hangsúlyhoz hozzájáruló akusztikai jellemzőt, az f_0 -t, az intenzitást és az időtartamot vizsgáltuk. Annak érdekében, hogy a szó jelentése ne befolyásolja a hangsúlyítéleteket, jelentés nélküli álszavakat használtunk. Feltételeztük, hogy a magyar nyelvben ezek a jellemzők a szakirodalmi, elsősorban angol és holland nyelvvel végzett tanulmányokban kapott eredményektől eltérően határozzák meg a hangsúlyítéleteket: feltételeztük az f_0 kiemelkedő fontosságát, illetve az időtartam kisebb jelentőségét.

Módszer

Résztvevők

A kísérletben 19 személy vett részt (15 nő, 4 férfi; átlagéletkor: 21,11 év, szórás: 1,76 év). A résztvevők önként vettek részt a kutatásban. Mindegyik résztvevő tanult legalább egy idegen nyelvet (4 egyet, 11 kettőt, 4 hármát), ezek közül a legtöbben angolt tanultak (17 személy). Minden résztvevő ép hallással rendelkezett. A végső elemzésből 4 személy adatait hagytuk ki, ennek okáról lásd az Eredmények részt. A kísérletet az Egyesített Pszichológiai Kutatásaitikai Bizottság engedélyezte.

Nyelvi anyag

A hangsúly percepciójának vizsgálatához szintetizált álszavakból álló nyelvi anyagot hoztunk létre. Ehhez a Profivox (Olaszy et al. 2000) beszéd-szintetizáló programot használtuk. A program lehetőséget biztosított az akusztikai paraméterek lehető legpontosabb megadására. A nyelvi anyagot egyetlen álszó, a Sluijter és munkatársai (1997) által is alkalmazott *nana* különféle akusztikai paraméterének változtatásával hoztuk létre. A hangsúly három akusztikai jellemzőjét változtattuk szisztematikusan (alapfrekvencia, intenzitás és időtartam) az első, illetve a második szótag esetében. A változtatás minden álszó esetében egyszerre csak egy szótagot és egy akusztikai jellemzőt érintett. 15 különböző álszót hoztunk létre változtatott paraméterenként, így a bemutatott nyelvi anyag összesen 45 álszóból állt (1. táblázat).

1. táblázat: Az álszavak akusztikai jellemzőinek értékei a változtatott akusztikai jellemzők függvényében

Álszó	f_0 (Hz)		Intenzitás (dB)		Időtartam (ms)	
	Első szótag	Második szótag	Első szótag	Második szótag	Első szótag	Második szótag
1.	157	132	81,2	67,4	210	160
2.	153	132	79,3	67,4	204	160
3.	149	132	77,3	67,4	197	160
4.	146	132	75,3	67,4	190	160
5.	142	132	73,4	67,4	183	160
6.	138	132	71,3	67,4	175	160
7.	135	132	69,3	67,4	167	160
8.	132	132	67,4	67,4	160	160
9.	132	136	67,4	69,2	160	167
10.	132	140	67,4	71,2	160	174
11.	132	143	67,4	73,3	160	183
12.	132	147	67,4	75,2	160	190
13.	132	151	67,4	77,2	160	197
14.	132	155	67,4	79,2	160	205
15.	132	159	67,4	81,2	160	210

Az 1–7. álszavak esetében az adott akusztikai jellemző nagyságát lépésenként csökkentettük az első szótagon, miközben a többi jellemző és a második szótag akusztikai paraméterei állandóak maradtak. A 8. álszónál nem volt különbség az első és második szótag között, a 9–15. álszó esetében pedig az adott akusztikai jellemzőt lépésenként növeltük a második szótagon. A semleges (8.) álszó akusztikai jellemzői a következők voltak: az alapfrekvencia 132 Hz, az intenzitás 67 dB, az időtartam pedig 160 ms volt mindkét szótagon. A változtatások nagyságát az alapfrekvencia és intenzitás esetében egy korábbi hangsúlyproduktions kísérlet adataira építve határoztuk meg, az idő-

tartamot pedig a beszédsszintetizáló program által előre megszabott legkisebb értékkel növeltük. Az akusztikai paraméterek lépésenkénti változtatásával perceptuális szinten az adott szótag hangsúlyosságát kívántuk változtatni: az 1. álszó esetében az első és második szótag közötti nagy akusztikai különbség az első szótag erőteljes hangsúlyosságát jelezte, ami egyre csökkent a 8. álszóig, ahol elvileg az első és második szótag hangsúlyossága megegyezett. Ezt követően a második szótag lett egyre hangsúlyosabb, egészen a 15. álszóig, ahol a legnagyobb volt a különbség a két szótag között a második szótag javára.

Eljárás

A kísérlet megkezdése előtt minden résztvevő tájékoztatásban részesült a kísérlet menetéről írásban és szóban is. Minden résztvevő kitöltött egy rövid kérdőívet a nemére, korára és kezességére vonatkozóan, illetve aláírt egy, a kísérletben való önkéntes részvételről szóló és az adatok anonimitását garantáló nyilatkozatot.

A kísérletet Presentation 12.1 (www.neurobs.com) programmal számítógépen futtattuk le. A résztvevők fülhallgatón keresztül hallgatták meg a nyelvi anyagot hangszigetelt laborban az MTA Természettudományi Kutatóközpont Kognitív Idegtudományi és Pszichológiai Intézetében. Válaszaikat egy billentyűzeten a jobb, illetve a bal nyíl gomb lenyomásával jelezték.

Minden álszó ötször szerepelt, így összesen 225 álszóról kellett diszkriminációs döntést hozni. A bemutatás véletlenszerűen történt, függetlenül az akusztikai jellemzőkben történt változtatásoktól. A résztvevők feladata az volt, hogy egy kétalternatívás kényszerválasztásos helyzetben jelezzék, hogy a bemutatott beszédinger első vagy második szótagja hallatszott számukra hangsúlyosabbnak. A kísérlet során adott válaszok mellett a döntés meghozatalához szükséges reakcióidők is rögzítésre kerültek. Az éles próbákat 5 próbából álló gyakorló fázis előzte meg.

Statisztikai elemzés

A reakcióidő-adatok elemzésének első lépésében kiszűrtük a kiugró értékeket, kizárva azokat az adatokat, amik túl alacsony vagy túl magas értékkel rendelkeztek. Azokat az értékeket tekintettük kiugrónak, melyek az összes adat átlagánál három szórással nagyobb, illetve kisebb értékkel rendelkeztek. Ezáltal a felső határt 3737 ms-nál, az alsó határt pedig 500 ms-nál határoztuk meg, mivel az ennél rövidebb idő alatti válaszokról úgy véltük, hogy nem megbízhatóak, tekintettel arra, hogy az álszavak átlagos időtartama 347 ms volt, és a megbízható döntéshez mindenképpen meg kellett hallgatni az álszó második szótagját is. Az ezek alapján kiszűrt adatokhoz tartozó kategorizációs válaszokat szintén kiszűrtük.

A kategorizációs döntés válaszai esetén személyenként kiszámoltuk, hogy a résztvevők az egyes álszavaknál az öt bemutatásból milyen arányban észlelték a hangsúlyt a második szótagon. A statisztikai elemzéseket ezen az arányszámon végeztük, és az ábrákon $p(2.szótag)$ elnevezéssel jelöltük. A reakció-

idők adatai esetén személyenként átlagoltuk az öt bemutatás reakcióidőit, és összesítettük őket.

A reakcióidőkön és a $p(2.szótag)$ értéken 3×15 -ös elrendezésű, Típus (időtartam, f_0 , intenzitás) \times Álszó (15 lépés) ismételt méréses varianciaanalízist (ANOVA) végeztünk. Greenhouse–Geisser-korrekciót alkalmaztunk abban az esetben, ha a szfericitás előfeltétele sérült (l. Vargha 2008: 362–363), és Tukey-féle HSD-tesztet alkalmaztunk a specifikus különbségek felderítésére. Az adatokat a STATISTICA 12 programmal elemeztük.

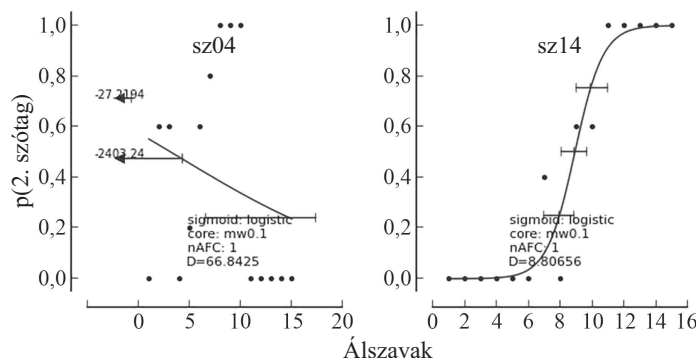
A kategorizációs döntések esetében megvizsgáltuk, hogy hogyan alakultak az egyéni hangsúlyítételek a három különböző akusztikai jellemző nagyságának függvényében. Mivel az akusztikai jellemzők értékei az első 7 lépésben folyamatosan csökkentek az álszó első szótagján, majd a második 7 lépésben folyamatosan nőttek a második szótagon, az eredményeket ideális esetben egy logisztikus görbével lehet ábrázolni, amennyiben azt mérjük, hogy mekkora arányban választották a résztvevők a második szótagra eső hangsúlyítéletet. Ennek megfelelően az egyéni adatokra egy logisztikus görbét illesztünk a Psignifit programcsomag segítségével (Fründ et al. 2011). Az illesztés során logisztikus szigmoid görbét használtunk (*mw-core* paraméterezéssel, l. Fründ et al. 2011), amely két paraméter becslését teszi lehetővé: az *m* paraméter azt a küszöbértéket méri, amelynél a válaszok gyakorisága meghaladja az 50%-ot (jelen esetben a második hangsúlyos válaszok gyakoriságát); a *w* paraméter az emelkedés meredekségét méri. Ezeken kívül a programcsomag két másik paramétert is megad: a *lambda*-t, amely a görbe felső aszimptotája és a kihagyott válaszok mérőszáma, valamint a *gamma*-t, amely a görbe alsó aszimptotája és a találgatás mérőszáma. Az illesztés során bootstrap módszerrel alapuló maximum likelihood módszerrel történő paraméterbecslést használtunk (Wichmann–Hill 2001). A görbék illeszkedésének jóság (goodness-of-fit) mutatójaként a deviance (*D*) értékét számoltuk ki. A kapott paraméterekben az egyes akusztikai jellemzők közötti különbséget a Típus (időtartam, f_0 , intenzitás) faktoron elvégzett ANOVA-val elemeztük.

Eredmények

Egyéni eredmények elemzése, adattisztítás

A kapott eredmények statisztikai elemzése előtt megvizsgáltuk az egyéni válaszok jellemzőit, és ezek alapján kiszűrtük azokat a személyeket, akik feltehetően nem megfelelően oldották meg a feladatot figyelmetlenség, a feladat félreértése vagy egyéb okok miatt. Ehhez az alábbi kritériumokat használtuk: a) reakcióidő-kritériumok (< 500 ms, > 3737 ms) miatt kiszűrt vagy a lehetséges válaszgomboktól eltérő gomb megnyomása miatti nagyszámú nem értékelhető válasz ($> 20\%$); b) egyéni adatokra illesztett pszichometria függvény nem megfelelő illeszkedése a görbe alakja és az illeszkedési paraméterek (*m*, *w*, *lambda*, *gamma*) alapján.

Ezen kritériumok alapján 4 résztvevőt zártunk ki a további elemzésekből. Az 1. ábra illusztrálja az egyik kizárt személy és egy véletlenszerűen kiválasztott személy válaszait. Jól látható, hogy a kizárt személy válaszai egyáltalán nem követik az elvárható szigmoid jellegű válaszmintázatot.



1. ábra

Pszichometria függvény az f_0 által meghatározott kategorizációs döntésben egy kizárt (sz4) és egy nem kizárt (sz14) személy esetében (Az x tengely az álszavak sorszámát, az y tengely a második szótagra adott hangsúlyítéletek arányát tünteti fel.)

Trochaikus torzítás

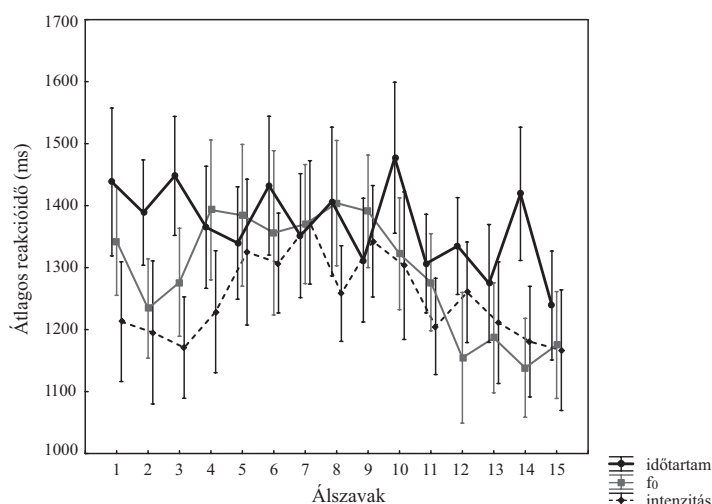
A vizsgálat során összesen 3375 választ adtak a résztvevők (45 álszó $\times 5$ ismétlés $\times 15$ résztvevő). A kiugró reakcióidők szűrése után 3313 válasz maradt. Ebből 2070 esetben az első szótagon lévő hangsúllyal rendelkező szót jelölték meg, ami a válaszok 62,5%-a. Ez az érték szignifikánsan a véletlen szint felett van (binomiális próba: $p < 0,001$).

Reakcióidő és kategorizációs döntés

A reakcióidő eredményeinek elemzésére (2. ábra) ismételt mérés varianciaanalízist alkalmaztunk két faktorral: 3 hangsúlyjellemző (időtartam, f_0 és intenzitás) és 15 különböző álszó.

A reakcióidő esetében szignifikáns Típus főhatás [$F(2,28) = 6,94$, $\varepsilon = 0,82$, $p < 0,01$, $\eta_p^2 = 0,33$] és tendenciaszinten Álszó főhatás [$F(14,196) = 2,23$, $\varepsilon = 0,29$, $p = 0,075$, $\eta_p^2 = 0,14$] jelentkezett. Az időtartamban változtatott álszavakra általában hosszabb volt a reakcióidő, mint az f_0 -ban és intenzitásban változtatott álszavakra, de ez a különbség a Tukey-féle HSD-teszt szerint csak az időtartam és intenzitás között volt szignifikáns ($p < 0,01$), míg az időtartam és az f_0 között csak tendenciaszinten jelent meg ($p < 0,1$). A reakcióidő-adatok azt mutatták, hogy a leghosszabb ideig azok a döntések tartottak, amelyek ese-

tében a legkisebb volt a különbség az első és második szótagon megjelenő akusztikai jellemzők nagysága között, vagyis a 8. lépés környékén.



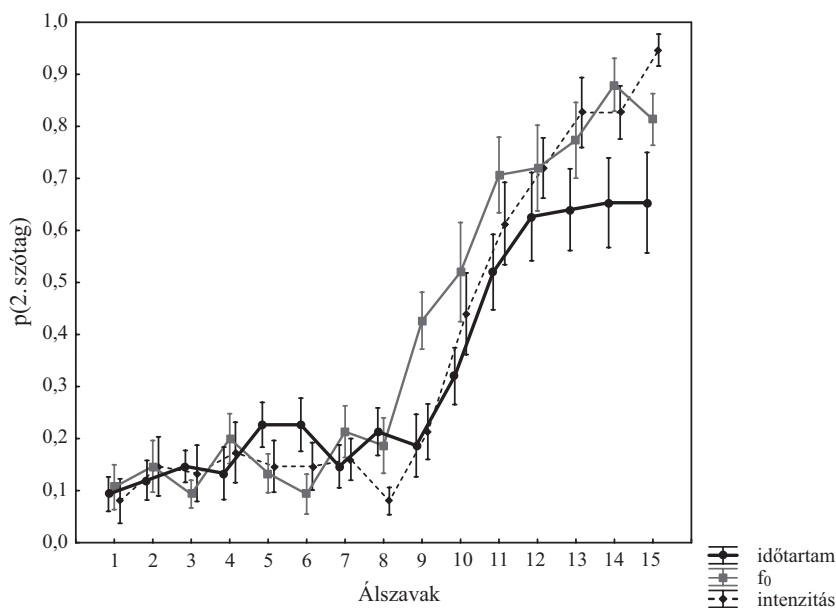
2. ábra

Átlagos reakcióidők (ms) az egyes álszavakra, változtatástól függően
(A hibavonalak a standard hibát jelzik.)

A kategorizációs döntés esetében az adatok alapján kiszámoltuk, hogy az álszavak ötszöri bemutatása során milyen arányban jelezték a résztvevők, hogy a hangsúlyt a második szótagon észlelték [$p(2. \text{ szótag})$]. A 3. ábrán látható, hogy milyen arányban estek a második szótagra a hangsúlyítételek a 15 álszó esetében, külön-külön a három akusztikai jellemzőre (időtartam, f_0 , intenzitás) lebontva. Az eredmények szerint a 9. álszóig a hangsúlyítételek nagy része (kb. 80%-a) az első szótagra esett. Ezután egy éles váltás látható a hangsúlyítételekben, és a válaszok fokozatosan átváltak a második szótagra. A kapott eredményeket két módon elemeztük statisztikailag. Egyrészt varianciaanalízisben összehasonlítottuk, hogy a 15 álszóra adott hangsúlyítételek valószínűsége eltér-e egymástól a három akusztikai jellemző mentén. Másrészt logisztikus függvényeket illesztettünk az egyéni hangsúlyítételekre, és azt vizsgáltuk, hogy a függvények paraméterei eltérnek-e egymástól a három akusztikai jellemző esetében.

Elsőként tehát a reakcióidő-adatokhoz hasonlóan ismételt mérésekre varianciaanalízist végeztünk, Típus (3 – időtartam, f_0 és intenzitás) \times álszó (15) faktorokkal. Szignifikáns Álszó főhatást kaptunk [$F(14, 196) = 52,91$, $\varepsilon = 0,21$, $p < 0,001$, $\eta_p^2 = 0,79$] és a Típus \times Álszó interakció is megjelent [$F(28, 392) =$

2,24, $\varepsilon = 0,28$, $p < 0,05$, $\eta_p^2 = 0,14$]. A Típus főhatás csak tendenciaszinten jelentkezett [$F(2, 28) = 3,14$, $\varepsilon = 0,88$, $p = 0,067$, $\eta_p^2 = 0,18$].



3. ábra

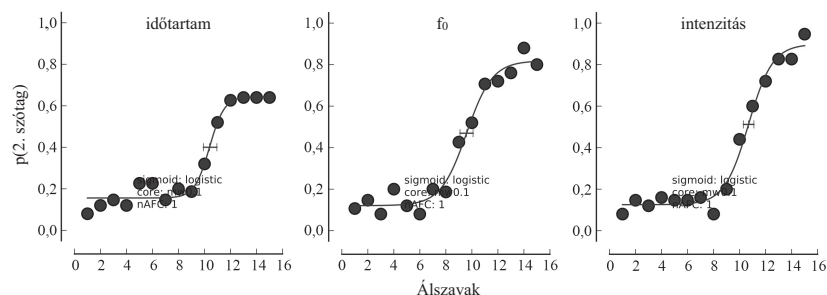
A résztvevők „második szótag” döntéseinek százaléka a különböző álszavakra, változtatástól függően (A hibavonalak a standard hibát jelzik.)

Az időtartamban változtatott álszavakra összességében kevesebb „második szótag” választ kaptunk, mint az f_0 -ban és az intenzitásban változtatott álszavakra. A Tukey-féle HSD-teszt alapján azonban csak tendenciaszintű különbség volt kimutatható az időtartamban változtatott álszavakra adott válaszok és az f_0 -ban változtatott álszavakra adott válaszok között ($p < 0,1$).

Pszichometria i függvény illesztése

A kategorizációs döntés esetében megvizsgáltuk, hogy hogyan alakultak az egyéni hangsúlyítételek a három különböző akusztikai jellemző nagyságának függvényében. A logisztikus függvényillesztés eredményét az összes résztvevő válaszáinak átlagán az időtartam, az f_0 és az intenzitás jellemzők esetében a 4. ábra mutatja be. Látható, hogy az adatpontokhoz jól illeszkedik a logisztikus görbe, és közel hasonlóak a küszöb- és meredekségjellemzők, ugyanakkor az időtartam-jellemzőre illesztett görbe aszimptotája kisebb, mint az f_0 és az intenzitás jellemzőké.

Az egyéni kategorizációs ítéletekre illesztett logisztikus függvények alapján kiszámolt paraméterértékek átlagát a három jellemző esetében a 2. táblázat mutatja be. A D értékek mindhárom jellemző esetében szignifikánsan jó illeszkedést jeleztek. A táblázatból látható, hogy a paraméterek értéke nagyjából hasonló volt mindhárom akusztikai jellemző esetében. Egyedül a λ értékében volt kismértékű eltérés az időtartam és a másik két jellemző között.



4. ábra

Pszichometria függvények a három akusztikai jellemző által meghatározott kategorizációs döntésben a teljes mintán

(Az x tengely az álszavak sorszámát, az y tengely a második szótagra adott hangsúlyítéletek arányát tünteti fel.)

Annak érdekében, hogy ezeket az eltéréseket statisztikailag is jellemezni tudjuk, kiszámoltuk az egyéni adatokra illesztett görbék paramétereit (m , w , λ , γ), és ezeket ismételt mérések varianciaanalízisben elemeztük. A statisztikai elemzés egyik paraméter értékében sem mutatott szignifikáns eltérést a Típus faktor szintje között ($p > 0,1$).

2. táblázat: Átlagos paraméterértékek (zárójelben a szórásértékek), az akusztikai jellemzők változtatásától függően

Paraméter	Időtartam	f_0	Intenzitás
m	9,97 (2,98)	10,12 (1,82)	10,79 (1,17)
w	4,74 (5,52)	5,14 (6,98)	3,69 (3,17)
λ	0,24 (0,29)	0,11 (0,17)	0,07 (0,13)
γ	0,09 (0,09)	0,11 (0,09)	0,12 (0,11)
D	13,33 (4,64)	11,93 (5,65)	12,98 (5,46)

Következtetések

Vizsgálatunkban a magyar hangsúlyt meghatározó akusztikai jellemzők percepcióját tanulmányoztuk egy kényszerválasztásos feladatban. Eredmé-

nyeink szerint mindhárom vizsgált jellemző (időtartam, f_0 , intenzitás) hozzájárult a hangsúly észleléséhez.

Az első szótagon egyre csökkenő, majd a második szótagon egyre növekvő akusztikai jellemzők értékei a kategoriális észlelésre jellemző szigmoid alakú logisztikus görbét eredményeztek, vagyis az első szótagra adott ítéletek hirtelen kezdtek csökkenni, majd váltottak át a második szótagra adott ítéletekre. A három akusztikai jellemző esetében nem találtunk statisztikailag alátámasztható különbséget abban, hogy a hangsúlyítéletek mikor és milyen meredekséggel váltanak át az egyik szótagból a másodikba. Mindhárom jellemző esetében igaz volt ugyanakkor az, hogy az 50%-os küszöbértéket a 10-11. álszónál érték el, vagyis jelentősen később, mint ahogy az akusztikai jellemzők áttevődtek a második szótagra (a 8. álszónál voltak azonos nagyságúak a jellemzők, és a 9. álszótól volt nagyobb a második szótagon ezek értéke). Ez arra utal, hogy a résztvevők tovább fenntartották az első szótagra adott hangsúlyítéleteket. Ezt az eredményt támasztja alá az is, hogy az összes válasz 62,5%-a „első szótag” ítélet volt.

A hangsúlyítéletekben tehát látható az erőteljes trochaikus torzítás, vagyis az első szótagra eső hangsúly preferenciája, abban az esetben is, amikor már az akusztikai jellemzők egyértelműen a második szótagon voltak nagyobbak. A Sluijter és munkatársai (1997) által holland hallgatókkal végzett vizsgálatban kapott 57%-os értékhez képest saját eredményeink nagyobb trochaikus torzítást jeleznek a magyar hallgatók esetében, ami nem meglepő, hiszen a magyar nyelvben kivétel nélkül minden szó esetében az első szótagra esik a hangsúly. Ezzel szemben a holland nyelvben az első szótagra eső hangsúly csak gyakoribb, mint a második szótagra eső hangsúly, de nem kizárólagos. Saját korábbi tanulmányunkban ezt a különbséget az első és második szótagra eső hangsúlymintázat között a legális/illegális vs. reguláris/irreguláris fogalmakkal írtuk le (Honbolygó–Csépe 2013), kiemelve ezzel, hogy a magyarban a második szótagra eső hangsúly szabályt sért, míg más szabad hangsúlyozású nyelvekben csak gyakoriságot. Ugyanakkor a magyar spontán beszédben megjelenhet a hangsúlyeltolódás (Gósy 2002), amelynek során a hangsúly az első szótagról áttevődhet egy másik szótagra az összetett szavak vagy bármely más szó esetében. Ez a jelenség azonban elsősorban a beszélt nyelvet érinti; Bóna és Imre (2007) eredményei szerint a hallgatók a legtöbb esetben nem veszik észre azt, miközben szövegértési teljesítményüket befolyásolja.

A hangsúlyítéletek reakcióidős és címkézési eredményei különbséget mutattak az időtartam, illetve az f_0 és az intenzitás jellemzők között: az időtartamban változtatott álszavakra hosszabb reakcióidőket és kevesebb második szótag ítéletet kaptunk, mint az f_0 -ban és intenzitásban változtatott álszavakra. Eszerint az időtartam információval jelzett hangsúllyal kapcsolatban a résztvevők lassabban hoztak hangsúlyítéletet, és a második szótagra való eltolódását a hangsúly elmozdulásának kevésbé erőteljes jelzéseként észlelték.

Az időtartam másik két akusztikai jellemzőtől való eltérését alátámasztja, hogy a kategorizációs döntésekre illesztett pszichometria függvény *lambda* értéke nem szignifikánsan ugyan, de kissé nagyobb volt, mint az f_0 és az intenzitás esetében.

Mindezen eredmények együttesen azt jelzik, hogy az időtartam információ kevésbé hatékony jelzése a hangsúlynak, mint az f_0 és az intenzitás. Ez nyilvánvalóan ellentmond az angol, illetve holland nyelvvel kapcsolatos szakirodalmi adatoknak (Fry 1958; Turk–Sawusch 1996; Sluiter et al. 1997), amelyek szerint a hangsúly legfontosabb jelzése az időtartam és az intenzitás. Ugyanakkor eredményeink azt is mutatják, hogy az időtartam a magyar nyelvben is felhasználható a hangsúly jelzésére, annak ellenére, hogy a magyarban az időtartam nem csak prozódiai, hanem szegmentális fonológiai jellemző is.

Nem találtunk különbséget az f_0 és intenzitás jellemzők által jelzett hangsúly feldolgozásában. A szakirodalmi adatok alapján az f_0 kiemelkedő fontosságát, illetve minden vagy semmi jellegű (Fry 1958) feldolgozását vártuk a többi jellemzőhöz képest. Eredményeink szerint mind az f_0 , mind az intenzitás feldolgozása minden vagy semmi jellegű volt, hiszen a logisztikus görbe mindkettő esetében hasonlóan meredeken emelkedett, és a hangsúlyítéletek nem fokozatosan, hanem hirtelen változtak meg, amikor az akusztikai jellemzők a második szótagon elkezdtek nagyobbá válni az elsőhöz képest. Ez kategoriális jellegű észlelésre utal, vagyis arra, hogy az akusztikai jellemzők esetében nem azok abszolút nagysága számít, hanem az, hogy diszkriminálható módon megjelenjenek az első vagy a második szótagon.

További vizsgálatokban érdemes lenne tanulmányozni a hangsúly olyan komplex akusztikai korrelátumainak a percepció jellemzőit is, mint a spektrális egyensúly. Valamint érdekes kérdés, hogy a hangsúly feldolgozásával kapcsolatos agyi folyamatok hogyan jelzik az akusztikai jellemzők változását, és mutatkozik-e eltérés ezekben a különböző jellemzők feldolgozása esetén.

Irodalom

- Bloomfield, Leonard 1993. *Language*. Holt, Rinehart and Winston, New York.
- Bóna Judit – Imre Angéla 2007. A hangsúlyeltolódás hatása a beszédfeldolgozásra. *Beszédkutatás* 2007. 75–82.
- Campbell, Nick – Beckman, Mary 1997. Stress, prominence and spectral tilt. *Intonation: Theory, Models and Applications. Proceedings*. Athens 18-20 September 1997. http://www.isca-speech.org/archive_open/archive_papers/int_97/inta_067.pdf. (A letöltés ideje: 2014. október 20.)
- Cutler, Anne – Norris, Dennis 1998. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 14. 113–121.
- Fónagy Iván 1958. *A hangsúlyról*. Nyelvtudományi értekezések 18. Akadémiai Kiadó, Budapest.

- Fründ, Ingo – Haenel, N. Valentin – Wichmann, Felix A. 2011. Inference for psychometric functions in the presence of nonstationary behavior. *Journal of Vision* 11/6. 1–19.
- Fry, Dennis Butler 1958. Experiments in the perception of stress. *Language and Speech* 1/2. 126–152.
- Gósy Mária 2002. A hangsúlyeltolódás jelensége. In Balázs Géza – A. Jászó Anna – Koltói Ádám (szerk.): *Éltető anyanyelvünk. Mai nyelvűvelésünk elmélete és gyakorlata. Írások Grétsy László 70. születésnapjára*. Tinta Könyvkiadó, Budapest. 193–198.
- Gósy Mária 2004. *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest.
- Kager, René 2007. Feet and metrical stress. In de Lacy, Paul (ed.): *The Cambridge handbook of phonology*. Cambridge University Press, Cambridge. 195–228.
- Honbolygó, Ferenc – Csépe, Valéria 2013. Saliency or template? ERP evidence for long-term representation of word stress. *International Journal of Psychophysiology* 87/2. 165–172.
- Olaszy Gábor – Kiss Géza – Németh Géza – Olaszi Péter 2000. Profivox: A legkorszerűbb hazai beszédszintetizátor és szövegfeldolvasó. *Beszédkutatás 2000*. 167–179.
- Plag, Ingo – Kunter, Gero – Schramm, Mareile 2011. Acoustic correlates of primary and secondary stress in North American English. *Journal of Phonetics* 39/3. 362–374.
- Sluijter, Agaath M. C. – van Heuven, Vincent J. 1996. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100/4. 2471–2485.
- Sluijter, Agaath M. C. – van Heuven, Vincent J. – Pacilly, Jos J. A. 1997. Spectral balance as a cue in perception of linguistic stress. *Journal of the Acoustical Society of America* 101/1. 503–513.
- Sweet, Henry 1906. *A primer of phonetics*. 3rd edition. Clarendon Press, Oxford.
- Turk, Alice – Sawusch, James R. 1996. The processing of duration and intensity cues to prominence. *Journal of the Acoustical Society of America* 99/6. 3782–3790.
- Vargha András 2008. Matematikai statisztika pszichológiai, nyelvészeti és biológiai alkalmazásokkal. Pólya Kiadó, Budapest.
- Wichmann, Felix A. – Hill, Jeremy 2001. The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics* 63/8. 1314–1329.

A kutatást az OTKA PD 84009 számú pályázata támogatta. Köszönjük prof. Olaszy Gábornak és prof. Németh Gézának (BME Távközlési és Médiainformatikai Tanszék), hogy a Profivox programot kutatási céllal rendelkezésünkre bocsátották.

A MEGNYILATKOZÁS VÉGÉNEK JELZÉSE FELOLVASÁSBAN: TEMPORÁLIS SZERKEZET ÉS ZÖNGEMINŐSÉG

Kohári Anna – Markó Alexandra

Bevezetés

A nagyobb prozódiai egységek határai kitüntetett szerephez jutnak a beszédprodukción és a beszédfeldolgozásban is, ezért a beszélők igyekeznek (nem feltétlenül tudatosan) valamilyen módon jelölni ezeket. A hangzó megnyilatkozások szerkezetének egyik elsődleges velejárója az egységek határán megjelenő temporális mintázat, amelynek a legjellemzőbb megnyilvánulása a határon megjelenő nyúlás (szuprasegmentális nézőpontból tempólassulás) (l. Fletcher 2010 áttekintését). Az irreguláris zöngeminőség (glottalizáció) ugyancsak gyakran tölt be határjelző funkciót, mind a frázis elején, mind a végén (l. Markó 2013 áttekintő összefoglalását). A jelen kutatásban ezt a két jelenséget vizsgáljuk magyar nyelvű felolvasásokban, a megnyilatkozások végén. Elemezzük, hogy milyen arányban jelentkezik irreguláris, illetve nem modális zöngeminőség (a leheletes zöngét is beleértve), valamint mennyire jellemző és milyen mértékű a lassulás magyar felolvasott mondatokban. Azt is megvizsgáljuk, hogy a két jelenség előfordulása között kimutatható-e valamilyen összefüggés.

Az intonációs frázis és/vagy megnyilatkozás végén jelentkező nyúlást sokan univerzális jelenségnek tartják, bár a mértéke és a kiterjedése (az, hogy hány szegmentumot érint) nyelvenként változó. Fletcher (2010) összefoglalásában egyebek között felsorolja az orosz, az angol több változatát, a franciát, az olaszt, a görögöt, a csehet, a németet, a hollandot, a svédet, a finnt, a japánt, az arabot, a hébert és a mandarint mint olyan nyelveket, amelyekben kimutatták ezt a jelenséget (a primer szakirodalmi forrásokat l. a hivatkozott helyen). A frázis végi nyújtás több nyelvben interakciót mutat például a hangsúlyozással, illetve a magánhangzó-minőséggel és a beszédtempó beszélfüggő variabilitásával.

A jelenség magyarázatára több elképzelést találunk a szakirodalomban. Lindblom (1975, idézi Fletcher 2010) például úgy véli, hogy a beszéd soron következő elemei egy hipotetikus „frázispuffer”-ben tárolódnak, és a frázis végi nyúlás egy általános lassulási tendencia eredménye, amelyet az idéz elő, hogy a pufferből egyre fogynak a beszéd elemei, majd a puffer kiürül. Más megközelítések (pl. Turk–Shattuck-Hufnagel 2000) szerint a frázis végi lassítás valójában egy hallgatóorientált stratégia az elemek különféle szintű össze-

tartozásának jelzésére. Öhman (1967) az artikulációs gesztusok renyhülésével magyarázza a jelenséget, és ezt sugallják Tabain (2003) artikulációs fonetikai kutatásának eredményei is.

Arra vonatkozóan ellentmondók az eredmények, hogy a megnyilatkozás végi nyúlás jelentősebb-e, mint az egyéb frázisvégeken mérhető időtartamtöbblet; illetőleg nyúlás/tempólassulás természetesen nemcsak a nagyobb prozódiai egységek határán jelenhet meg, hanem máshol is (vö. pl. Fletcher 2010). Ezért ezeknek a kérdéseknek a megválaszolásához komplexebb, a diszkurzus szerkezet és a prozódiai struktúra közötti kapcsolatot többértően figyelembe vevő elemzésekre volna szükség. Mivel a kutatók többsége azt feltételezte, hogy a nyúlás csak a frázis utolsó szótagján jelentkezik, kevés vizsgálat vonatkozott a megelőző szótagokra, és ezek eredményei között is ellentmondások látszanak.

A nyúlás mértéke több kutatás kérdéseként szerepelt. A franciában jelentős (akár 200-300%-os) időtartam-különbséget is mértek a megnyilatkozás végén és az ez előtt megjelenő magánhangzók között (Tabain 2003), míg az amerikai angolban átlagosan csak 90%-nyival hosszabbnak mérték a frázis végén ejtett szótagokat a korábbiaknál (Turk–Shattuck–Hufnagel 2007). Ugyanakkor természetesen meg kell jegyezni, hogy a hangsúlyozási mintázatok, illetve az ezekkel együtt járó magánhangzó-realizációs sajátosságok (pl. szűkségekppen rövidebb időtartamú svá a hangsúlytalan szótagban) nyelvenként különböző módon befolyásolják az eredményeket. Mindemellett nemcsak a magánhangzók, hanem a mássalhangzók is nyúlhatnak a megnyilatkozás végén (vö. pl. Oller 1973; Nakai et al. 2009).

A korábbi magyar nyelvű kutatások szerint a felolvasásokban mind a magánhangzók, mind a mássalhangzók hosszabbak abszolút hangsor végi helyzetben, mint közlés közben (Magdics 1966; Kassai 1979). A hangsor végi lassítást tényként kezelik a gépi felolvasó rendszerek modelljeiben is (vö. Olasz 2006; Tóth 2013).

A magyarban a megnyilatkozás végi nyúlásra vonatkozóan White és Mádý (2008) végzett szisztematikus kutatást. Minimálpár mondatokat hoztak létre egy, két és három szótagos, *o*-t és *ó*-t kváziazonos fonetikai kontextusban tartalmazó szavakkal, figyelembe véve a fókuszhangsúly mondatbeli helyzetét is. Eredményeik szerint a megnyilatkozás végi nyúlás a magyarban egyértelműen létezik, annak ellenére, hogy a kvantitás megkülönböztető szereppel bír a magánhangzók körében. Ez a fonológiai különbség a nyúlás megvalósulásában is nyomot hagy: a rövid magánhangzók nyúlása csak a megnyilatkozás utolsó szótagjában volt megfigyelhető, míg a hosszú magánhangzók esetében az utolsó előtti szótagban is időtartamtöbblet volt mérhető. Ezzel szemben Kovács (2002) vizsgálatai alapján a rövid magánhangzókon is kimutatható az időtartambeli növekedés az utolsó előtti szótagban. (Az idézett két kutatás módszere eltérő volt.)

A magyar beszédben az intonációs frázisokon belül a szavak tendenciaszerűen lassulnak az egységek vége felé (Váradí–Beke 2013). Az egymást követő magánhangzós és mássalhangzós szakaszok időtartamának vizsgálatai is azt mutatták, hogy lassuló tendencia figyelhető meg a tagmondatnyi egységeken belül (Kohári 2013). Kohári (2014) a beszéd időbeli változásait az egymást követő beszédhangok és szakaszok időtartamainak különbsége felől közelítette meg egy, a beszédhangokra adaptált leíró statisztikai módszer segítségével, és egyebek mellett azt találta, hogy a beszédszakaszok rendkívül változatosan realizálódtak ugyan, de jellemzően mégis lassuló trendet mutattak.

Az irreguláris zöngeminőség (glottalizáció) frázis/megnyilatkozás végi határjelző szerepét több nemzetközi kutatás igazolta. A szakirodalom (elsősorban Slifka 2007) alapján feltételezhető, hogy a glottalizáció határjelző funkciójának fiziológiai háttere is van: a zöngé felépülése és lecsengése nagyobb mértékben idézi elő a zöngé irreguláritását a közlés egységei kezdetén és végén. Henton és Bladon (1988) megnyilatkozás végi helyzetben mutatta ki a brit angol presztízsváltozatában; illetve az amerikai angolban az irreguláris hangszalagrezgés szintén a megnyilatkozás végét jelző akusztikai kulcsok egyike (Slifka 2006). Svéd olvasott beszéd frázishatárain is megjelenik az irreguláris zöngé (Fant–Kruckenberg 1989), akárcsak a finnben, a csehben és a „szerbhorvát”-ban (Lehiste 1965, idézi Gordon–Ladefoged 2001). Prozódiai egységek kezdetén és végén a szlovénban is gyakori (Jurgec é. n.).

A glottalizáció határjelző funkcióját magyar felolvasott és spontán szövegekben több különböző méretű beszédegység szintjén vizsgálták: beszédszakaszokban, intonációs frázisokban, megnyilatkozásokban és társalgási egységekben (Markó 2013, 2014). A beszédszakaszok (szünetekkel körülhatárolt szövegegységek) és az intonációs frázisok határain nem volt kimagasló arányú az irreguláris zöngé jelenléte: sem a vizsgált szövegegységek elején, sem végén nem volt gyakoribb a glottalizáció, mint az egységek belsejében. (Ez arra enged következtetni, hogy a fent említett fiziológiai motiváció hatása itt nem jelentős.) Ezzel szemben felolvasott szöveg mondatvégein (korábbi kutatások – Böhm–Ujváry 2008; Markó 2010 – eredményeivel is egybehangzóan) gyakran jelentkezik glottalizáció, akárcsak a spontán beszéd megnyilatkozásainak végén. A glottalizációs határjelzés kifejezetten a szerkezet végén (ahhoz közeledve, az utolsó néhány szótagon) jelent meg, és olykor áthúzódtott a következő egység elejére. A frázis eleji glottalizáció ugyanakkor a magyarban nem annyira jellemző, mint más nyelvekben (l. fent), kivéve a V#V határokat (vö. Markó 2013). A BEA szövegfelolvasásaiban a glottalizáció szempontjából azt figyeltük meg, hogy a beszélők jellemzően a megnyilatkozások utolsó öt szótagján glottalizálnak (vö. Markó 2013). A szakirodalom is megerősíti, hogy a megnyilatkozás végi glottalizáció sokszor nemcsak a legutolsó szótagon jelentkezik. Sőt azoknak az eseteknek a száma sem elhanyagolható, amikor a glottalizáció nem a megnyilatkozás utolsó szótagján, hanem az azt megelőző két-három szótagon jelentkezik, és az utolsó szótagra

regulárisra változik a zöngé minősége. Ezeket az eseteket is megnyilatkozás végi glottalizációnak tekintik a szerzők (pl. Henton–Bladon 1988; Slifka 2006; Böhm–Ujváry 2008).

Az irreguláris zöngé megjelenési gyakorisága erősen beszélőfüggő (l. pl. Henton–Bladon 1988; Böhm–Ujváry 2008; Markó 2013). Az eredmények szerint ugyanakkor a megnyilatkozathatár glottalizációval való megjelölése még azoknál a beszélőknél is relatíve gyakori, akiknek a zöngéképzése egyébként ritkán vált át irregulárisba. Sőt kimondható, hogy minél kevesebbet glottalizál egy beszélő, annál valószínűbb, hogy ezt megnyilatkozást záró frázishatáron teszi. (Kivételt képeznek ez alól a rádióbemondók, akik kifejezetten kerülnek ezt a megoldást, feltehetően beszédtanári instrukció alapján, vö. Markó 2013.)

A nem modális zöngeminőség másik viszonylag gyakori formája egyes nyelvekben a leheletes zöngé (vö. pl. Kohler 2000; Ishi et al. 2010), amelynek sajátossága, hogy zöngéképzés közben a hangszalagok között kiszökik a levegő, ezáltal turbulens zaj keletkezik, amelynek a lenyomata „zörejjel kevert zöngéként” látszik a regisztrátumon (Gordon–Ladefoged 2001). Egy korábbi kutatás eredménye alapján az is felmerült, hogy a glottalizáció és a leheletes zöngé funkciókörében lehet átfedés (l. ugyancsak Markó 2013), illetőleg nemzetközi kutatások alapján egyes esetekben kizáró disztribúcióban állnak egymással (Kohler 2000). Ezért indokoltnak tartottuk a két jelenség együttes elemzését is.

A jelen kutatásban felolvasások megnyilatkozásainak záró részét vizsgáljuk a fent bemutatott két jellemző alapján: a temporális szerkezetet (összevetve a megnyilatkozás többi részének temporális szerkezetével), valamint az irreguláris, illetve a nem modális (irreguláris és/vagy leheletes) zöngeminőség megjelenését. Hipotézisünk szerint a glottalizáció megjelenése összefügg a lassítások megjelenésével. Egy részről elképzelhető, hogy a két jelenség jellemzően együtt fordul elő közösen jelölve a megnyilatkozásvéget. Ugyanakkor ennek ellenkezője is feltételezhető, hogy kiegészítő viszonyban vannak egymással, tehát a beszélők azokban az esetekben, amikor glottalizációval jelzik a határ közeledését, nem/kevésbé alkalmaznak lassítást a megnyilatkozás végéhez közeledve; és fordítva: tempólassítás alkalmazásakor a zöngéképzésük nem/kevésbé alakul irregulárisba. Feltételezzük továbbá, hogy a különálló mondatok felolvasása esetén mindkét jelenség (a tempólassítás és a glottalizáció is) gyakoribb, illetve mértékük nagyobb a mondat végéhez közeledve, mint a szövegfelolvasás mondataiban. Ennek háttérében azt feltételezzük, hogy a különálló mondatok önállóbb egységek, mint a szövegmondatok, és ezért ezek a megnyilatkozásvégek nagyobb mértékben lehetnek jelöltek.

Kísérleti személyek, anyag, módszer

A kutatáshoz a BEA adatbázis (Gósy et al. 2012) tíz beszélőjének mondat- és szövegolvasását használtuk fel. Az adatközlők fele nő, fele férfi, nem dohá-

nyoznak, nincs zöngképzési vagy más beszédzavaruk, és nincs ismert hallásproblémájuk. Életkoruk 20 és 60 év között szóródik, az átlagéletkor 41,3 év.

A mondatfelolvasás a BEA protokollja szerint az utolsó előtti feladat: itt azokat a mondatokat kell felolvasniuk az adatközlőknek, amelyeket a felvétel kezdetén hallás után megismételtek. A 25 kijelentő mondat szószáma 5 és 10 között alakul, átlagosan 8,1 szó/mondat. A szövegolvasás a felvétel utolsó része, az adatközlőknek egy tudományos ismeretterjesztő szöveget kell rövid előzetes felkészülés után felolvasniuk. A szöveg 12 kijelentő mondatból (és a címből) áll, a mondatok szószáma 8 és 33 között szóródik (a címet nem számítva), átlagosan 19,2 szó/mondat.

A korábbi kutatás (vö. Markó 2013) alapján 12 beszédhangban határoztuk meg a megnyilatkozás végét, minden felolvasott mondat esetében az utolsó 12 elhangzott beszédhangot elemeztük. A tíz beszélőnek 374 olyan megnyilatkozása (246 önálló mondat és 128 szövegmondat) volt, amelynek utolsó 12 hangjában nem fordult elő nyelvbottlás, félreolvasás. Így összesen 4488 beszédhangot elemeztünk zöngeminőségük és időtartamuk szempontjából. A lassuló vagy gyorsuló trend megállapításához nemcsak a mondatok utolsó 12 hangját vizsgáltuk meg, hanem összevetésként az azokat megelőző 12 hangot is, így további 4488 hang időtartamát vontuk be a vizsgálatba.

Ezt megelőzően természetesen hangszinten annotáltuk a hanganyagot a magyar nyelvre is adaptált MAUS elnevezésű automatikusan szegmentáló szoftverrel (Schiel 1999). Jelöltük a beszédhangok minőségét, határait, valamint a szüneteket. Az így kapott hanghatárokat manuálisan ellenőriztük a Praat 5.3 szoftverben (Boersma–Weenink 2013). A magánhangzók időtartamát a formánsszerkezet kezdetéhez és végéhez igazítottuk az oscillogram, a spektrogram és az auditív információk segítségével, követve a nemzetközi szakirodalomban szokásos hangelhatárolási kritériumokat (Grabe–Low 2002). Azokban az esetekben, ahol a hanghatárok nem voltak egyértelműek az oscillogram és a spektrogram alapján (pl. magánhangzó áll nazális mássalhangzó vagy approximáns környezetében), a hanghatárt a formánsátmenet felénél helyeztük el. A szünetet követően megjelenő zöngétlen felpattanó zárhangok zárszakaszának idejét úgy határoztuk meg, hogy ugyanazon közlés következő, ugyanolyan képzésű hangzó zárszakaszának idejét hozzáadtuk a zárhang zörejes részéhez. Ha az utolsó 12 beszédhang kiejtése közben a beszélő szünetet tartott, a szünetet és annak időtartamát figyelmen kívül hagytuk.

A Praatban megjelenítettük a hullámformát és a hangszínképet; illetőleg folyamatosan és többször ellenőrizve hallgattuk le a hanganyagokat. A hangszinten annotált hanganyagon beszédhangonként címkéztük a zöngeminőséget a következőképpen: külön jelöltük, ha a hangzó zöngeminősége nem vizsgálható (zöngétlen obstruens), illetve ha nem ítéltető meg – ide tartozott az esetleges zaj (a felvételek végéhez közeledve papírzörgés), valamint egyésgesen ide soroltuk a zöngés obstruenseket és a pergőhang realizációit, mivel az akusztikus kép alapján ezeknél nem mindig dönthető el a zöngeminő-

ség. A glottalizációra lehetőséget adó szegmentumok (magánhangzók és szonor mássalhangzók) esetében jelöltük, ha a zöngé modális, ha irreguláris, illetve ha leheletes volt. Ha az adott beszédhang bármely kis részletében irreguláris periódusokat tapasztaltunk, a hangot glottalizáltnak címkéztük. A glottalizált realizációk annotálása a korábbi kutatások módszertanához igazodva (pl. Dilley et al. 1996; Böhm–Ujváry 2008) a vizuális és auditív információk együttes figyelembevételével történt. Akusztikai szempontból glottalizáltnak tartottuk az adott beszédhangrészletet, ha az alapperiódusok időtartama vagy amplitúdója hirtelen jelentősen megváltozott; vagy ha hirtelen a beszélő normál/szokásos hangterjedelme alá csökkent az alaphfrekvencia. Mindemellett percepció szempontként tekintetbe vettük, hogy a szegmentum hangszínezete jól hallhatóan érdes, rekedtes. Akkor címkéztük glottalizáltnak a beszédhangrészletet, ha az akusztikai lenyomaton látható, és ezzel egyidejűleg auditív úton is észlelhető volt az irregularitás. A leheletes zöngé jelölési kritériuma szintén az volt, hogy a kiszökő levegő zöreje a zöngelenyomatban látható, valamint hallható legyen.

Fontos megjegyezni, hogy a glottalizáció többfunkciós jelenség. Az a tény tehát, hogy egy adott beszédhang irreguláris zöngével realizálódott, nem feltétlenül jelenti azt, hogy az adott szegmentumon a glottalizációnak az a szerepe, hogy a megnyilatkozás végének közeledtét jelezze. Ennek megfelelően csak akkor címkéztünk glottalizáltnak egy szegmentumot, ha az biztosan nem két szomszédos magánhangzó határát és/vagy fráziskezdetet jelölt (vö. Markó 2013).

Mindezek alapján a megnyilatkozásvégek beszédhangjai körében a zöngeminőséget elemezve két körben vizsgáltunk. Az egyik esetben a 12 vizsgált beszédhang közül meghatároztuk azoknak a beszédhangoknak a számát, amelyeknek a zöngeminősége vizsgálható (és ezt mérni is tudtuk, azaz a magánhangzók és a szonor mássalhangzók tartoztak ide). Ezeknek a beszédhangoknak a számát tekintettük 100%-nak, és azt elemeztük, hogy ezek milyen arányban realizálódtak ténylegesen glottalizáltan. A másik esetben az utolsó 12 hang közül csak a magánhangzókat vettük tekintetbe (az összes magánhangzó száma volt 100%), és azt vizsgáltuk, hogy ezek közül mennyi volt glottalizált. Nemcsak az utolsó 12 hang glottalizáltságának arányát vizsgáltuk meg, hanem külön az utolsó 6 és külön az utolsó 3 hangét is. Mindkét megközelítésben végeztünk olyan összesítést is, ahol a leheletes zöngeminőséget összevontuk az irregulárisal, vagyis a nem modális zöngét együttesen jelöltnek vettük.

A temporális szerkezetet a beszédritmus-mérőszámokat alkalmazó szakirodalomhoz hasonlóan a hangzók időtartamából kiindulva vizsgáltuk meg (Grabe–Low 2002). A lassulás és a gyorsulás méréséhez az egyik beszédritmus-mérőszám (PVI) és az ún. lépéssztatisztikai módszer alapttechnikáját használtuk fel. Ennek lényege, hogy a vizsgált egység időtartamát összehasonlítjuk a követő hangzó időtartamával, ily módon lehetővé válik a folyama-

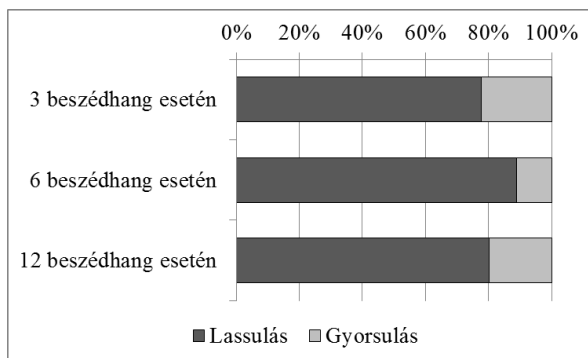
tos beszédben lévő hangsor temporális viszonyainak feltárása. Az adott hang időtartamából kivonjuk a követő hang időtartamát, a különbség előjele megadja, hogy a két egység közötti viszony lassuló vagy gyorsuló. Ha a követő hang időtartama nagyobb, lassulásról beszélünk, ha rövidebb, akkor gyorsulásról. A különbségek mértéke pedig megadja a lassulás vagy gyorsulás mértékét. A módszer kiterjeszthető nagyobb egységek vizsgálatára, és összehasonlítására is, így a több hangon átívelő jelenségek mérése is lehetővé válik. A különböző egységek összehasonlításához azonban nem a beszédhangok időtartamát, hanem az adott egységben lévő hangok darabszámát osztottuk el a beszédhangok teljes időtartamával, azaz az artikulációs tempót vettük figyelembe, ezzel is csökkentve a szélsőséges időtartamértékek hatását. Az utolsó 12 hangra kapott artikulációs tempót kivontuk az azt megelőző 12 hangra kapott artikulációs tempó értékéből. A különbség előjele alapján eldönthető, hogy a megnyilatkozás vége lassult vagy gyorsult. A lassulások és gyorsulások elemzéséhez azonban nem állapítható meg optimális ablakméret (l. de Looze 2010), ezért különböző nagyságú egységeket is megvizsgáltunk a glottalizációhoz hasonlóan. Összehasonlítottuk az utolsó 3 és 6 beszédhangra számolt artikulációs tempót is az azt megelőző, ugyanakkora egységek artikulációs tempójával, amelyeket szintén kategorizáltunk az alapján, hogy lassulásnak vagy gyorsulásnak tekinthetők. Az összetett időzítési mintázatok ezen módszer vizsgálatából származó eredményei nem kezelhetők szigorú kategóriákként, hiszen például az egyes beszédhangok ún. intrinzikus időtartama eltérő (vö. pl. Gósy 2004; Olaszy 2006), de az időtartamok viszonyainak ily módon történő leegyszerűsítése nagyobb trendek, összefüggések kimutatására alkalmas lehet (l. Kohári 2013). A különféle egységekre kapott esetleges azonos trendmegjelölések pedig mutathatják az eredmények megbízhatóságát is.

Elemeztük a fenti paramétereket a teljes vizsgálati korpusz tekintetében, valamint megvizsgáltuk a mondat- és a szövegfelolvasás közötti lehetséges eltéréseket is. A statisztikai elemzéseket (χ^2 -próba, Mann–Whitney-próba, Spearman-féle korrelációelemzés) az SPSS 20.0 program segítségével végeztük el.

Eredmények

A megnyilatkozások végét általánosságban csökkenő artikulációs tempó, azaz lassulás jellemzi. Amennyiben az utolsó 3 beszédhang artikulációs tempóját viszonyítottuk az azt megelőző 3 hang artikulációs tempójához, akkor az összes megnyilatkozás 77,8%-ában a megnyilatkozás utolsó szakaszának artikulációs tempója kisebb volt, mint a megelőző szakaszé. Amennyiben az utolsó 6 beszédhang és az azt megelőző 6 hang artikulációs tempóját vettük figyelembe, ez az arány valamelyest nagyobbak bizonyult (89,0%). A gyorsuló trendet mutató szakaszvégek aránya mindkét esetben egyértelműen alacsonynak mutatkozott (3 beszédhang esetében: 22,2%, 6 beszédhang eseté-

ben: 11,0%). A megnyilatkozások végének lassulása még az utolsó 12 beszédhang és az azt megelőző 12 hang összehasonlításakor is egyértelműen tetten érhető. Az utolsó szakaszok artikulációs tempója az esetek 80,3%-ában alacsonyabb a megelőző szakasz artikulációs tempójánál. A megnyilatkozás végi lassulás tehát a szakirodalommal megegyezően következesen megjelenik a megnyilatkozások végén, ugyanakkor a lassulás nemcsak az utolsó egy-két szótagnyi nagyságú egységeket jellemzi, hanem nagyobb szövegrészekben is felfedezhető (1. ábra). A különböző nagyságú egységek összehasonlítása azonban csak az általános trendet jelzik, tehát nem a szakaszban lévő összes hang nyúlásáról van szó. A lassuló trendet mutató esetek artikulációs tempókülönbségének átlagos értéke kisebb is a 12-12 beszédhangot összehasonlító elemzésnél: 12 hang esetében 2,9 hang/s (szórás: 1,4 hang/s); 6 beszédhang esetében 3,9 hang/s (szórás: 2,0 hang/s); 3 beszédhang esetében 3,6 hang/s (szórás: 2,2 hang/s).

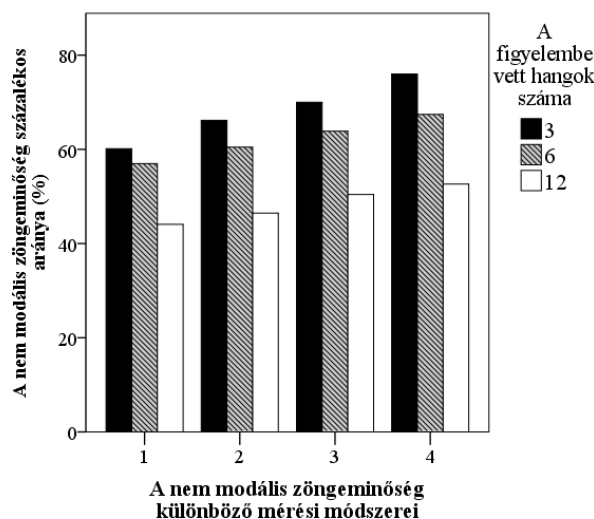


1. ábra

A lassuló és gyorsuló szakaszvégek aránya különböző méretű vizsgált egységek esetén

Az irreguláris zöng megjelenése ugyancsak jellemző a megnyilatkozásvégeken, és a közlés végéhez közeledve egyre nagyobb a glottalizált beszédhangok aránya. A 2. ábra azt szemlélteti, hogy az összes megnyilatkozás különböző méretű szakaszvégeiben átlagosan hány százalékos a glottalizáció/irreguláris zöngképzés megjelenése. (A 100% minden esetben azoknak az adott típusú szegmentumoknak a számát jelenti, amelyek az adott egységben egyáltalán irreguláris zöngével valósulhattak meg).

Az ábrán egyértelműen látszik az a tendencia, hogy a megnyilatkozás végéhez közeledve mind önmagában az irregularitás, mind együttesen a modálistól eltérő zöngeminőség egyre nagyobb arányú. Továbbá ha a vizsgálatot a magánhangzókra korlátozzuk, ezek az arányok még nagyobbak, mint az összes szonor ($V + C_{szon}$) körében.



2. ábra

A jelöltség (irregularitás, illetve a modálistól való eltérés) mértéke a különböző nagyságú egységek és az elemzési tartományok függvényében (1 = az irreguláris zöngével megvalósult szegmentumok aránya az összes $V + C_{szon}$ számához viszonyítva; 2 = a nem modális zöngével megvalósult szegmentumok aránya az összes $V + C_{szon}$ számához viszonyítva; 3 = az irreguláris zöngével megvalósult V-k aránya az összes V számához viszonyítva; 4 = a nem modális zöngével megvalósult V-k aránya az összes V számához viszonyítva)

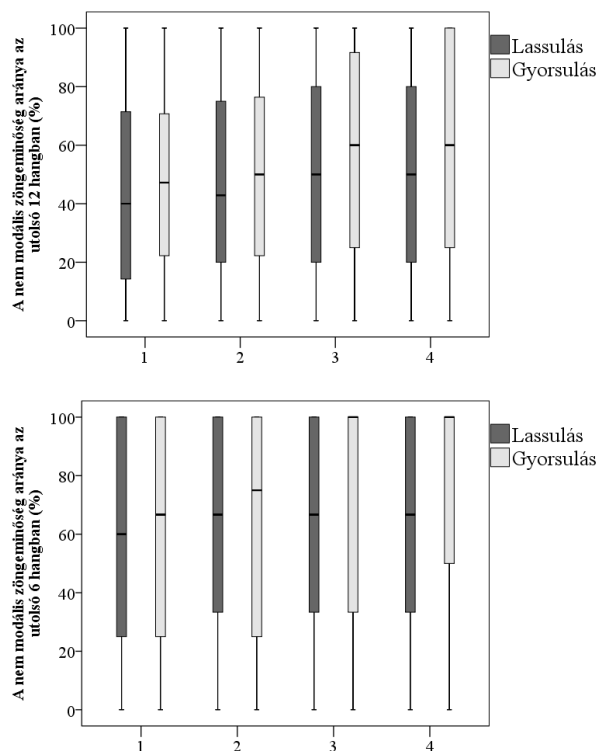
A 12 beszédhangra terjedő elemzés eredménye szerint az összes szonor beszédhang közül a glottalizáltak aránya 44,1% (szórás: 31,0%), míg az összes nem modális realizáció aránya 46,5% (szórás: 31,1%). A magánhangzók között az irregulárisan ejtettek aránya 50,4% (szórás: 33,7%), a nem modális realizációké 52,6% (szórás: 33,9%). Ha az elemzést 6 beszédhangnyi terjedelembre korlátozzuk, az összes szonor ($V + C_{szon}$) 57,0%-a (szórás: 35,9%) irreguláris, és összesen 60,5%-a (szórás: 35,3%) nem modális. Ebben a körben a magánhangzók 63,9%-a (szórás: 38,2%) glottalizált, 67,4%-a (szórás: 37,5%) nem modális. Ha csak a megnyilatkozások utolsó 3 beszédhangját tekintjük, a szonorok ($V + C_{szon}$) 60,1%-a (szórás: 42,2%) irreguláris és 66,2%-a (szórás: 40,0%) nem modális; míg a magánhangzók 70,0%-a (szórás: 45,0%) glottalizált, és 76,0%-a (szórás: 42,0%) nem modális. Mindebből az is kiderül, hogy a leheletes zöng a megnyilatkozás végén viszonylag ritka, míg a glottalizáció igencsak gyakori. Ugyanakkor a megnyilatkozás utolsó 6 beszédhangján (vagyis a 2-3 záró szótagon) mindkét jelenség gyakrabban fordul elő, mint korábban.

Ahogy vártuk, mind a glottalizáció (illetve tágabban a nem modális zöngeminőség), mind a lassulás gyakori jelenség a megnyilatkozás végén. Felmerül a kérdés, hogy csak a megnyilatkozás vége okozza a két paraméter együttes előfordulását, vagy összefüggés is kimutatható a megjelenésük között. Összevetettük a lassuló és a gyorsuló trendet mutató megnyilatkozásvégeket a glottalizált és a nem modális zöngeminőség megjelenési aránya szerint. Azt tapasztaltuk, hogy akár az utolsó 12 hangot, akár az utolsó 6 hangot vizsgáltuk, a glottalizáltság vagy a modálistól eltérő zöngeminőség aránya a gyorsuló szakaszokban volt nagyobb (3. ábra). A szórások azonban minden esetben nagynak mondhatók, a statisztikai próbák nem is támasztották alá az eltérések szignifikáns voltát (12 beszédhang esetében a Mann–Whitney-próba: $Z = 11399,5$, $p \geq 0,091$; 6 beszédhang esetében pedig: $Z = 7743,0$, $p \geq 0,312$). Ez azt jelenti, hogy ezekben az egységekben sem a lassulás és a glottalizáció, sem a gyorsulás és a glottalizáció következetes együtt járása nem igazolható statisztikailag.

A jelöltség mértékét és a temporális sajátosságot az utolsó 3 hang esetében is összevetettük. Mivel az utolsó 3 hang vizsgálatakor a szonor mássalhangzókra számolt glottalizáltság vagy modálistól való eltérés aránya 5-féle kategóriát vehet fel, a magánhangzókra számolt pedig csak 3-féleképpen realizálódott, ezért a megjelenési arány vizsgálatától eltekintettünk. Helyette azt vizsgáltuk meg, hogy az utolsó három beszédhang szonor mássalhangzói és/vagy magánhangzói eltérnek-e a modális zöngétől, ha lassuló, vagy ha gyorsuló trendet mutató szakaszcsoportról van szó (4. ábra). Akkor tekintettünk egy szakaszt glottalizáltnak vagy nem modálisnak, ha az utolsó három hang közül a magánhangzók vagy a szonor mássalhangzók valamelyike így realizálódott.

Amennyiben az összes szonort ($V + C_{\text{szon}}$) figyelembe vettük, azt tapasztaltuk, hogy a glottalizált, lassuló trendet mutató szakaszvégek 55,1%-ban fordulnak elő, míg a nem glottalizált, lassuló szakaszok 22,7%-ban. A megnyilatkozásvégek 17,1%-a glottalizált, gyorsuló trendet mutató szakaszként realizálódott. Alig volt olyan eset (5,1%), amelyben sem lassulás, sem glottalizáltság nem volt megfigyelhető a megnyilatkozás végén. A többi mérési módszer eredményeként is hasonló arányokat kaptunk. A megnyilatkozásvégek magánhangzói és szonor mássalhangzói 60,4%-ban nem modálisok voltak, 17,4%-ban modálisok lassuló trend esetében. A megnyilatkozások 18,2%-a volt nem modális, és csak 4,0%-a modális a gyorsuló szakaszokban. Amennyiben csak a magánhangzókat vettük figyelembe, a megnyilatkozásvégek 54,3%-a volt glottalizált a lassuló szakaszokban, 23,5%-uk volt nem glottalizált a lassuló szakaszokban, 16,0%-uk volt glottalizált a gyorsuló szakaszokban, és 6,1%-uk volt nem glottalizált a gyorsuló szakaszokban. Amennyiben a magánhangzóknál a lehetséges zöngét is figyelembe vettük, a megnyilatkozásvégek 59,1%-a volt nem modális a lassuló szakaszokban, 18,7%-uk volt modális a lassuló szakaszokban, 17,1%-uk volt nem modális a gyorsuló szakaszokban, végül 5,1%-uk volt modális a gyorsuló szakaszokban.

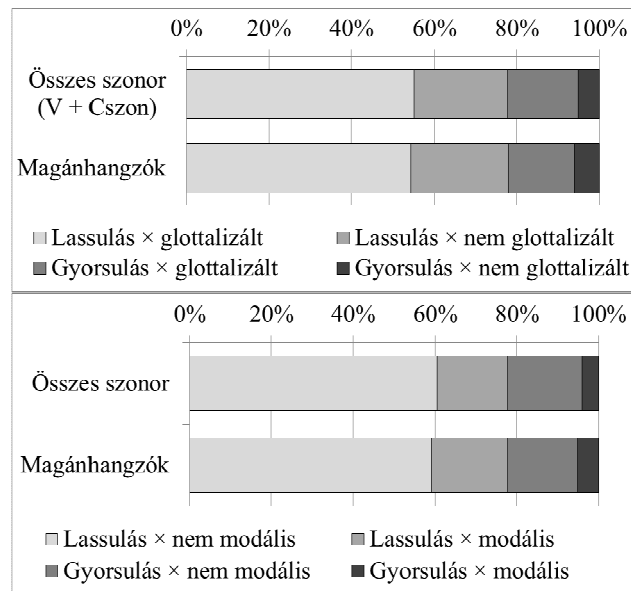
ban. Összességében tehát elmondható, hogy a megnyilatkozások végét vagy lassulás, vagy nem modális zöngeminőség jelzi, igen ritkán fordul elő, hogy egyik sem jelenik meg a megnyilatkozás végén. Ugyanakkor a két paraméter nem mutat összefüggést a χ^2 -próba alapján egyik, a zöngeminőség mérésére alkalmazott módszerünk esetében sem ($\chi^2 = 1,254$; $p > 0,257$), tehát a megnyilatkozás végi glottalizáció vagy modálistól eltérő zöngeminőség megjelenése függetlennek mutatkozik az artikulációs tempó változtatásától.



3. ábra

A jelöltség (irregularitás, illetve a modálistól való eltérés) mértéke a lassuló és a gyorsuló trendet mutató utolsó 12 hangban (fent) és 6 hangban (lent)

(1 = az irregularis zöngével megvalósult szegmentumok aránya az összes V + C_{szon} számához viszonyítva; 2 = a nem modális zöngével megvalósult szegmentumok aránya az összes V + C_{szon} számához viszonyítva; 3 = az irregularis zöngével megvalósult V-k aránya az összes V számához viszonyítva; 4 = a nem modális zöngével megvalósult V-k aránya az összes V számához viszonyítva)



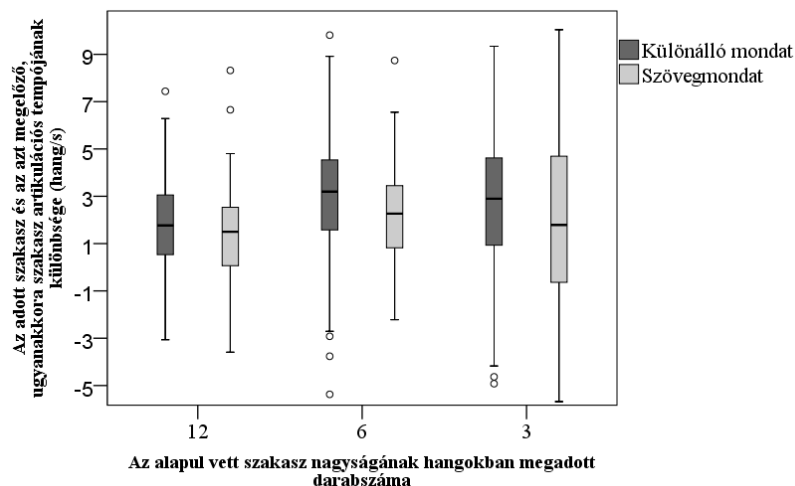
4. ábra

A temporális változások jellege és a zöngeminőség közötti összefüggések a megnyilatkozásokat záró 3 beszédhang esetén

Az eddigiekben a megnyilatkozás végi temporális változást statikus jelen-ségként kezeltük, a lassulásnak és a gyorsulásnak azonban a mértéke is meg-ragadható. A vizsgált szakaszok temporális változásának mértékét úgy hatá-roztuk meg, hogy a szakasz nagyságával azonos nagyságú megelőző szakasz artikulációs tempójából kivontuk a vizsgált szakasz artikulációs tempóját. Az így kapott különbséget tekintettük a szakaszra jellemző lassulás vagy gyorsu-lás jellemző mértékének. Az artikulációs tempó különbségének értéke nem mutatott összefüggést a glottalizált vagy a nem modális hangok arányával. Az utolsó 12 beszédhangot alapul véve a temporális változás mértéke nem mutatott szignifikáns összefüggést semelyik lehetséges zöngeminőséggel kapcsolatos aránnyal ($\rho \leq -0,066$, $p \geq 0,209$). Az utolsó 6 beszédhang eseté-ben szintén nem mutatott szignifikáns összefüggést a statisztikai próba ($\rho \leq -0,090$, $p \geq 0,084$). (3 beszédhangra ez az összefüggés nem volt vizsgálható.) A megnyilatkozások végén tehát a gyorsítás vagy a lassítás mértéke nem függ össze az ott megjelenő hangok zöngeminőségével.

Megvizsgáltuk a mondat- és a szövegfelolvasás közötti lehetséges eltéré-seket is. A megnyilatkozás végén lévő lassulás és gyorsulás mértéke függött attól, hogy a beszélők különálló mondatokat vagy szövegben szereplő mon-datokat olvastak fel. A szövegolvasásnál kisebb az artikulációs tempóban lé-

vő különbség (vagyis egyenletesebb az artikulációs tempó), mint különálló mondatok felolvasásakor, függetlenül az alapegység nagyságától. A megnyilatkozások utolsó 12 beszédhangjának és az azt megelőző 12 hang artikulációs tempójának különbsége a mondatfelolvasásokban átlagosan 1,8 hang/s (szórás: 1,9 hang/s), a szövegfelolvasásokban 1,3 hang/s (szórás: 2,0 hang/s). 6 beszédhangot alapul véve a mondatfelolvasásban mért artikulációs tempó különbsége átlagosan 3,2 hang/s (szórás: 2,6 hang/s) volt, a szövegfelolvasásban pedig 2,2 hang/s (szórás: 1,9 hang/s). 3-3 hang artikulációs tempóját összehasonlítva pedig a mondatfelolvasásokban átlagosan 2,8 hang/s (szórás: 2,7 hang/s), a szövegfelolvasásokban 1,7 hang/s (szórás: 3,5 hang/s) volt a különbség (5. ábra). A Mann–Whitney-próba mindhárom esetben szignifikánsnak mutatta az eredményeket (12 beszédhang esetében: $Z = 12676,0$; $p = 0,028$; 6 beszédhang esetében: $Z = 20752,0$; $p < 0,001$; 3 beszédhang esetében $Z = 21227,0$; $p < 0,001$).



5. ábra

A gyorsulás és a lassulás mértéke a felolvasás típusától függően a különböző nagyságú szövegegységekben

Felmerül a kérdés, hogy vajon a szövegfelolvasás és a mondatfelolvasás jellemző artikulációs tempója befolyásolhatta-e az eredményeket, ezért megvizsgáltuk a mondatok teljes hosszára számolt artikulációs tempót. Az artikulációs tempó azonban mindkét típusban nagyon hasonló (mondatfelolvasás esetében az átlag 13,1, a szórás 1,3 hang/s; szövegfelolvasás esetében az átlag 13,2, a szórás 1,1 hang/s). A két típus mondatainak artikulációs tempója nem tér el szignifikánsan ($Z = 14747,0$; $p = 0,315$), így a jellemző artikulációs

tempó nem befolyásolhatta a lassulás vagy gyorsulás mértéke és a felolvasás típusa között talált összefüggést.

A felolvasandó anyag eltérő volta nemcsak a temporális változások mértékére, hanem a glottalizációs arányokra is befolyással lehet, ezért megvizsgáltuk a glottalizált vs. nem glottalizált, illetve a modális vs. nem modális arányokat is a felolvasás típusának függvényében. A jelöltségi arányok átlagosan magasabbnak mutatkoztak a szövegben, mint a felolvasásokban. Az utolsó 12 beszédhang esetében a szonorok ($V + C_{szon}$) átlagosan 43,0%-ban (szórás: 30,4%) voltak glottalizáltak különálló mondatok felolvasása esetében, a szövegfelolvasásban pedig 46,1%-ban (szórás: 31,9%). Szintén 12 beszédhangnyi egységek esetében a szonorok ($V + C_{szon}$) átlagosan 45,1%-a (szórás: 30,4%) realizálódott nem modális zöngéképzéssel különálló mondatok felolvasásakor, a szövegfelolvasásban viszont 49,1%-uk (szórás: 32,4%). A magánhangzók glottalizáltsága szintén átlagosan nagyobb arányú volt szövegfelolvasásokban (átlag: 52,3%, szórás: 33,8%), mint különálló mondatokban (átlag: 49,4%, szórás: 33,8%). A magánhangzók modálistól eltérő volta is gyakrabban fordult elő szövegfelolvasásokban (55,2%, szórás: 34,3%), mint különálló mondatokban (átlag: 51,3%, szórás: 33,6%). A statisztikai próbák azonban egyik esetben sem támasztották alá, hogy a különbség jelentős lenne ($Z = 16933,5$; $p \geq 0,230$). 6 beszédhang esetében a szonorok ($V + C_{szon}$) átlagosan 54,8%-a (szórás: 35,3%) realizálódott glottalizáltan különálló mondatok felolvasásakor, a szövegfelolvasásban viszont 61,0%-uk (szórás: 36,9%). Ugyanekkora szakaszhosszúságnál a magánhangzók átlagos glottalizációs aránya 62,0% (szórás: 38,1%) volt különálló mondatok felolvasásában, 67,6% (szórás: 38,3%) pedig szövegfelolvasásban. A statisztikai próba itt sem mutatott szignifikáns eltérést ($Z = 17375,0$; $p \geq 0,095$). Ha azonban ugyanekkora nagyságú egységet vizsgálva a modálisok és a nem modálisok arányát elemeztük a felolvasás típusának függvényében, a statisztikai próbák igazolták az eltérések jelentőségét, amely eltérések a következőképpen alakultak. A szonorokra ($V + C_{szon}$) számolt nem modális arány kisebbnek mutatkozott különálló mondatokban (átlag: 58,0%, szórás: 34,6%), mint szövegfelolvasásokban (átlag: 65,2%, szórás: 36,3%), amely különbség szignifikáns ($Z = 17771,5$; $p = 0,031$). A magánhangzókra számolt arány is átlagosan nagyobbak mutatkoztak a szövegfelolvasásokban (átlag: 71,9%, szórás: 37,9%), mint a különálló mondatokban (átlag: 65,1%, szórás: 37,2%), és az eltérés itt is szignifikáns a statisztikai próba alapján ($Z = 17575,0$; $p = 0,048$). A nem modális zöngeminőség tehát statisztikailag is alátámaszthatóan gyakrabban jelent meg a szövegfelolvasásokban, mint a különálló mondatok felolvasásakor.

Következtetések

Kutatásunkban a megnyilatkozások végének két lehetséges jelölési módját, a nem modális zöngeminőséget (ezen belül az irreguláris és a leheletes zön-

gét), valamint a temporális szerkezetet, illetve ezek összefüggéseit vizsgáltuk magyar nyelvű felolvasásokban.

Az eredmények szerint a lassulások nagy arányban jelennek meg a megnyilatkozásvégeken a korábbi kutatási eredményekhez hasonlóan, ugyanakkor a lassulás nemcsak az utolsó egy-két szótagnyi nagyságú egységeket jellemzi, hanem az artikulációs tempó csökkenése nagyobb szakaszokban (akár 6, 12 hangnyi távolságban) is felfedezhető. Az irreguláris zöng megjelenése szintén gyakori jelenség a megnyilatkozásvégeken (Böhm–Ujváry 2008; Markó 2010 eredményeivel is egybehangzóan), a megnyilatkozás végéhez közeledve mind önmagában az irregularitás, mind együttesen a modálistól eltérő zöngeminőség egyre nagyobb arányú. Habár az eredményekből nem dönthető el egyértelműen, hogy mekkora szakasz vizsgálata lenne alkalmasabb a szakaszvégi lassulások és a glottalizáció megragadására, a megnyilatkozás utolsó 6 beszédhangján (vagyis a 2-3 záró szótagon) mindkét jelenség gyakrabban fordul elő, mint ezt megelőzően.

Mind a lassulás, mind a zöng nem modálissá válása jelzi a megnyilatkozás végét, de a két tényező között nem találtunk lineáris korrelációt, sem a gyakoriságuk, sem a mértékük tekintetében ebben a pozícióban. A lassulások és gyorsulások mérésére ugyanakkor nincs kialakult, bevált módszertan, tehát az általunk használt módszer csak egy lehetséges megközelítés. A jelen tanulmányban közölt méréseken túl ezért további lehetséges eljárásokat is megvizsgáltunk. Többek között az artikulációs tempó különbsége helyett a nyúlás százalékos értékét határoztuk meg (hasonlóan: Turk–Shattuck–Hufnagel 2007), illetve finomítottunk a bináris kategorizáción, azaz a lassulás és gyorsulás mellé alkottunk egy stagnáló csoportot is, amely csoportba azon megnyilatkozások végeinek adatai kerültek, ahol a változás értéke nem haladt meg egy általunk meghatározott (10% vagy 20%) küszöbértéket. Az ekképpen módosított módszertannal lefuttatott elemzések azonban nem hoztak a fent tárgyaltaktól eltérő eredményt, ezért nem is részleteztük őket.

Összegezve: úgy tűnik, hogy pusztán a megnyilatkozás vége idézi elő a két paraméter együttes előfordulását, és alig fordul elő olyan eset, hogy egyik jelenség sem jelenik meg megnyilatkozásvégen. A megnyilatkozás vége tehát valamelyik vagy mindkét tényező által nagy valószínűséggel jelöltté válik. Ez a megállapítás nyilvánvalóan csak az adott anyagra, tehát a felolvasásra érvényes, hogy más beszédmódok esetén hogyan érvényesül a határjelölés, további kutatásokat igényelne. Ugyanakkor az általunk vizsgált anyagban is találtunk eltéréseket a felolvasott anyag jellegétől függően. A különálló mondatok felolvasása esetében a beszélők jobban lassítottak a megnyilatkozás végén, mint a szövegfelolvasásokban. A glottalizáció esetében ugyan nem mutatkozott ilyen szisztematikus, statisztikailag is kimutatható eltérés a két felolvasástípus között, de a szövegmondatok végi glottalizáció valamelyest nagyobb mértékű volt. Ezek az eltérések véleményünk szerint két okra vezethetők vissza, ezek közül az egyik a beszélő attitűdje, szándéka, a kétféle

anyag felolvasása közben. A különálló mondatok végi lassítás mögött okként húzódhat az, hogy a különálló mondat lezárt, kompakt egység, ezért önmagában erősebben jelölt a megnyilatkozásvég. A nagyobb arányú glottalizáció a szövegmondatokban pedig talán éppen a folytatástól való markánsabb eltérés jelölésének szándéka miatt lehetséges. A másik lehetséges magyarázatot a mondatok hosszában látjuk: mivel a szöveg mondatai átlagosan jelentősen hosszabbak, mint a különálló mondatok. A rövidebb különálló mondatok esetében talán emiatt nagyobb mértékben kontrollálható a tempóváltoztatás, a hosszú szövegmondatokban pedig esetleg a hangterjedelemi lehetőségek szabnak határt az f_0 -változásnak, és ezért csap át gyakrabban irregulárisba a zönge.

Felvetődött, hogy a glottalizáció és a leheletes zönge funkciókörében lehet átfedés, a leheletes zönge kisszámú előfordulása miatt azonban ez a kérdéskör ebben az anyagban nem volt vizsgálható.

Összességében tehát nem sikerült közvetlen kapcsolatot kimutatni a nem modális zöngeminőség és a lassulás megjelenése között a megnyilatkozások végén, ennek ellenére az adatok arra utalnak, hogy a tempólassulás és a glottalizáció jellegzetes mintázatokat vesz fel, és ezek között vannak bizonyos összefüggések.

Irodalom

- Boersma, Paul – Weenink, David 2013. *Praat: doing phonetics by computer [Computer program]. Version 5.3.* <http://www.praat.org>. (A letöltés ideje: 2013. júl. 9.)
- Bóhm Tamás – Ujváry István 2008. Az irreguláris fonáció mint egyéni hangjellemző a magyar beszédben. *Beszédkutatás* 2008. 108–120.
- Dilley, Laura – Shattuck-Hufnagel, Stefanie – Ostendorf, Mari 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24. 423–444.
- Fant, Gunnar – Kruckenberg, Anita 1989. Preliminaries to the study of Swedish prose reading and reading style. *Speech Transmission Laboratory Quarterly Progress and Status Report* 30/2. Royal Institute of Technology, Stockholm. 1–80.
- Fletcher, Janet 2010. The prosody of speech: Timing and rhythm. In Hardcastle, William J. – Laver, John – Gibbon, Fiona E. (eds.): *The handbook of phonetic sciences*. 2nd edition. Wiley-Blackwell, Oxford. 521–602.
- Gordon, Matthew – Ladefoged, Peter 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics* 29. 383–406.
- Gósy Mária 2003. Virtuális mondatok a spontán beszédben. *Beszédkutatás* 2003. 19–43.
- Gósy Mária 2004. *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest.
- Gósy Mária – Gyarmathy Dorottya – Horváth Viktória – Grácz Tekla Etelka – Beke András – Neuberger Tilda – Nikléczy Péter 2012. BEA: beszélt nyelvi adatbázis. In Gósy Mária (szerk.): *Beszéd, adatbázis, kutatások*. Akadémiai Kiadó, Budapest. 9–24.

- Grabe, Esther – Low, Ee Ling 2002. Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology* 7. 515–546.
- Henton, Caroline – Bladon, Anthony 1988. Creak as a sociophonetic marker. In Hyman, Larry M. – Li, Charles N. (eds.): *Language, speech and mind. Studies in honour of Victoria A. Fromkin*. Routledge, London–New York. 3–29.
- Hunyadi, László 2002. *Hungarian sentence prosody and universal grammar. On the phonology – syntax interface*. Metalinguistica 13. Peter Lang, Frankfurt/M.–Berlin–Bern–Bruxelles–New York–Oxford–Wien.
- Ishi, Carlos Toshinori – Ishiguro, Hiroshi – Hagita, Norihiro 2010. Analysis of the roles and the dynamics of breathy and whispery voice qualities in dialogue speech. *EURASIP Journal on Audio, Speech, and Music Processing* 2010. <http://asmp.eurasipjournals.com/content/pdf/1687-4722-2010-528193.pdf>. (A letöltés ideje 2015. január 20.)
- Jurcec, Peter é. n. *Creaky voice in Slovene*. www.jurcec.net/publications/Creaky.pdf. (A letöltés ideje 2012. október 20.)
- Kassai Ilona 1979. *Időtartam és kvantitás a magyar nyelvben*. Nyelvtudományi Értekezések 112. Akadémiai Kiadó, Budapest.
- Kohári, Anna 2013. Temporal patterns of segments and intervals in Hungarian language. In Mertens, Piet – Simon, Anne Catherine (eds.): *Proceedings of the Prosody-Discourse Interface Conference (IDP)*. Leuven. 51–57. http://www.ling.arts.kuleuven.be/franitalco/idp2013/papers/Mertens_Simon_2013_Proceedings_IDP2013.pdf. (A letöltés ideje 2014. január 7.)
- Kohári Anna 2014. Tempóváltozások a vizsgált szakasz nagyságának függvényében. *Beszéd kutatás 2014*. 220–236.
- Kohler, Klaus J. 2000. Linguistic and paralinguistic functions of non-modal voice in connected speech. In: *Proceedings of the 5th Seminar on Speech Production: Models and Data*. Kloster Seeon, Germany. 121–124.
- Kovács Magdolna 2002. *Tendenciák és szabályszerűségek a magánhangzó-időtartamok produkciójában és percepciójában*. Debreceni Egyetem Kossuth Egyetemi Kiadója, Debrecen.
- Lehiste, Ilse 1965. Juncture. In *Proceedings of the 5th International Congress of Phonetic Sciences, Münster 1964*. S. Karger, New York. 172–200.
- Lindblom, Björn 1975. Some temporal regularities of spoken Swedish. In Fant, Gunnar – Tatham, M. A. A. (eds.): *Auditory analysis and perception of speech*. Academic Press, New York. 387–396.
- de Looze, Celine 2010. *ADoReVA and ADoTeVA, two PRAAT plugins for the Automatic Detection of Register and Tempo Variations*. http://celinedelooze.com/2010DeLooze_Submission_SPSASSD.pdf. (A letöltés ideje 2014. január 7.)
- Magdics Klára 1966. *A magyar beszédhangok időtartama*. Nyelvtudományi Közlemények 68. 125–139.
- Markó Alexandra 2010. A prozódia szerepe a spontán beszéd tagolásában. *Beszéd kutatás 2010*. 82–99.
- Markó Alexandra 2013. *Az irreguláris zöngé funkciói a magyar beszédben*. ELTE Eötvös Kiadó, Budapest.
- Markó Alexandra 2014. Az irreguláris zöngemínőség gyakorisága és pozíciói különféle spontán beszédhelyzetekben. *Beszéd kutatás 2014*. 69–86.

- Nakai, Satsuki – Kunnari, Sari – Turk, Alice – Suomi, kari – Ylitalo, Riikka 2009. Utterance-final lengthening and quantity in Northern Finnish. *Journal of Phonetics* 37. 29–45.
- Nespor, Marina – Vogel, Irene 1986. *Prosodic phonology*. Foris Publications, Dordrecht.
- Olaszy Gábor 2006. *Hangidőtartamok és időszerkezeti elemek a magyar beszédben*. Akadémiai Kiadó, Budapest.
- Oller, D. Kimbrough 1973. The effect of position in utterance on speech segment duration in English. *The Journal of the Acoustical Society of America* 54/5. 1235–1247.
- Öhman, Sven E. G. 1967. Numerical model of coarticulation. *Journal of the Acoustical Society of America* 41. 310–320.
- Schiel, Florian 1999. Automatic phonetic transcription of non-prompted speech. In Ohala, J. John – Hasegawa, Yoko – Ohala, Manjari – Granville, Daniel – Bailey, Ashlee C. (eds.): *Proceedings of the 14th International Congress of Phonetic Sciences*. University of California, San Francisco. 607–610.
- Slifka, Janet 2006. Some physiological correlates to regular and irregular phonation at the end of an utterance. *Journal of Voice* 20/2. 171–186.
- Slifka, Janet 2007. Irregular phonation and its preferred role as a cue to silence in phonological systems. In Trouvain, Jürgen – Barry, William J. (eds.): *Proceedings of the 16th International Congress of Phonetic Sciences. Saarbrücken 6–10 August 2007*. Pirrot GmbH., Dudweiler. 229–232.
- Tabain, Marija 2003. Effects of prosodic boundary on /aC/ sequences: Articulatory results. *Journal of the Acoustical Society of America* 113. 2834–2849.
- Tóth Bálint Pál 2013. *Rejtett Markov-modell alapú gépi beszédeltés*. Doktori értekezés. BME, Budapest. http://www.omikk.bme.hu/collections/phd/Villamosmernoki_es_Informatikai_Kar/2013/Toth_Balint_Pal/ertekezes.pdf. (A letöltés ideje 2013. dec. 4.)
- Turk, Alice E. – Shattuck-Hufnagel, Stefanie 2000. Word-boundary-related duration patterns in English. *Journal of Phonetics* 28. 397–440.
- Turk, Alice E. – Shattuck-Hufnagel, Stefanie 2007. Multiple targets of phrase-final lengthening in American English. *Journal of Phonetics* 35. 445–472.
- Váradi Viola 2008. A virtuális mondatok műfaji meghatározottsága. *Beszéd kutatás 2008*. 134–147.
- Váradi Viola 2009. Határjelzés a spontán beszédben és a felolvasásban. In Gecső Tamás – Sárdi Csilla (szerk.): *A kommunikáció nyelvészeti aspektusai*. Tinta Könyvkiadó, Budapest. 278–283.
- Váradi Viola – Beke András 2013. Az artikulációs tempó variabilitása felolvasásban. *Beszéd kutatás 2013*. 26–41.
- Varga László 1994. A hanglejtés. In Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 2. Fonológia*. Akadémiai Kiadó, Budapest, 468–549.
- Varga László 2002. *Intonation and stress. Evidence from Hungarian*. Palgrave Macmillan, Houndmills, Basingstoke.
- White, Laurence – Mády, Katalin 2008. The long and the short and the final: Phonological vowel length and prosodic timing in Hungarian. In: Barbosa, Plinio A. – Madureira, Sandra – Reis César (eds.): *Speech Prosody 2008. Fourth Conference on Speech Prosody. Proceedings*. Campinas, Brasil. 363–366.

MAGÁNHANGZÓK TEMPORÁLIS JELLEMZŐI AZ IDŐ MÚLÁSÁNAK FÜGGVÉNYÉBEN

Gósy Mária – Krepsz Valéria

Bevezetés

A hangképzés forrás-szűrő modelljében (Stevens 1999) a magánhangzók artikulációjakor a tüdőből kiáramló levegő megrezegteti a hangszalagokat, ez a hangforrás, majd módosul a toldalékcsőben, amely szűrőként funkcionál. A forrás és a szűrő is relatíve rugalmasan változik különféle tényezők érvényesülésének hatására. A hangképzésben bekövetkező változások mind ugyanazon beszélő esetében, mind különböző beszélők esetében következményesen megjelennek az akusztikai szerkezetben. A következmények tetten érhetők számos akusztikai-fonetikai paraméterben, ilyen például a teljes spektrum, a formánsértékek, a formánsávszélességek, a turbulens zörejek, a hangmagasság változása, avagy az időtartam. Mindezek a tényezők eltérőek az egyes beszélők között tekintettel a forrás és a szűrő egyéni sajátosságaira, illetve egyéni működtetésére (Zhang et al. 2006). Számos kutatás igazolta ugyanakkor, hogy ugyanazon beszélő esetében sem állandók a beszédhangok és a hangkapcsolatok akusztikai-fonetikai jellemzői, mivel a különböző tényezők hatására módosul az artikuláció, és ennek megfelelően változik az akusztikum (pl. McDougall 2006; Zhang et al. 2006; Beckford Wassink et al. 2007).

A hangképzést meghatározó tényezőket szokásosan két csoportba osztják; az egyikbe az ún. hosszú távú hangjellemzők, a másikba a rövid távú hangjellemzők tartoznak (pl. Mooshammer et al. 2008). Az előbbieket alapvetően az anatómiai sajátosságok határozzák meg, az utóbbiak pedig az adott nyelv vagy nyelvjárás hangképzési jellemzőin alapszanak. Egyszerűbben megfogalmazva, a hangképzés változatossága a beszélőszervektől és a beszélőszervek működtetésétől függ. Mindennek következtében a hangképzés eltérő lehet a különböző kommunikációs helyzetekben. A magánhangzók kiejtésének egyik relatíve változó paramétere a fonetikai időtartam. Az adott nyelvre jellemző fonológiai időtartam fonetikailag sokféle lehet; a fonológiai „hosszú” és „rövid” fizikai tartamok erősen változóak, és jellegzetes átfedéseket mutatnak (pl. Gósy–Beke 2010).

A magánhangzók ejtésére hatással vannak a zöngképzés jellemzői, a fonetikai kontextus, a koartikulációs jelenségek, a hangsúly, a beszéddallam, a fonotaktikai jellemzők, a közlés teljes hossza, a beszédstílus, a beszélő neme, érzelmei, szociolingvisztikai és pszicholingvisztikai faktorok, sőt a beszédhang gyakorisága is, és a felsorolás még folytatható (Harmegnies–Landercy

1988; Heuvel et al. 1996; Kovács 2002; Greenberg et al. 2003; Zhang et al. 2006; Recasens–Espinosa 2006; Benzeguiba et al. 2007; Dromey–Sanders 2009; Gósy–Beke 2010; da Silva et al. 2011; Kahn et al. 2011). Amerikai angol (férfi ejtésű) magánhangzók esetében kimutatták, hogy az időtartamukra hét tényező hat (de a tempót és a szintaktikai szerkezetet külön nem elemezték): a magánhangzó minősége, a szomszédos beszédhangok, a magánhangzó helyzete a szótagban, illetve a szótag típusa, a megelőző és a követő szótagok száma a mondatban, a szótaghangsúly, valamint a szó helyzete a mondatban (van Santen 1992). Természetesen a magánhangzók artikulációjának azért többé-kevésbé állandónak kell lennie a megfelelő észlelés érdekében, függetlenül az elvárt ejtés megközelítésének mértékétől és irányától (Lindblom 1990; Drager 2010).

Felmerül a kérdés, hogy mi történik a magánhangzók időtartamának alakulásával akkor, ha egy bizonyos idő eltelik az egyébként változatlan tartalmú közlések között. Pontosabban fogalmazva úgy tehetjük fel a kérdést, hogy milyen mértékben módosulnak a magánhangzók időviszonyai abban az esetben, ha a szokásosan befolyást gyakorló faktorok relatíve állandóak maradnak, csak valamennyi idő eltelik a közlések között. Vajon az idő múlása milyen hatást gyakorol a magánhangzók időtartamára ugyanazon beszélők és különböző beszélők esetében? Az artikulációs időkontroll képes-e ellensúlyozni azt a tényt, hogy a beszélő különböző időpontokban produkálja a magánhangzót? A jelen kutatásunk kérdése az, hogy az idő múlása milyen mértékben befolyásolja az egyébként csaknem változatlan körülmények között létrehozott közléseket (esetünkben mondatokat) és a közlések elemzett magánhangzóinak időtartamát.

Az „időt” mint az egyik kulcstényezőt a beszédprodukciónban leggyakrabban az életkor változásából adódó eltérések esetében vizsgálták (pl. Gayraud et al. 2011; Schötz et al. 2012). Magyar magánhangzók időtartamát és formánsait elemezték ugyanazon mondat ismétlései alapján 10 hónapon keresztül négy női beszélő ejtésében (Gósy 2002). Az eredmények jellegzetes beszélőn belüli és beszélők közötti eltéréseket igazoltak; az időtartamokat tekintve három beszélőnél volt szignifikáns a különbség az eltelt idő függvényében.

A jelen kutatásban négy magánhangzó fizikai időtartamát, a hordozó mondatok kiejtési idejét és artikulációs tempóját vizsgáltuk 7 hónapos periódusban, kéthetenkénti hangrögzítés alapján. Hipotéziseink a következők voltak: (i) a vizsgálati idő hét hónapja alatt a magánhangzók fizikai időtartamára az állandóság jellemzőbb lesz, mint a változás, (ii) az időbeli állandóságot a magánhangzó minősége befolyásolja, (iii) feltételeztük, hogy az ismételt mondatok artikulációs tempója nem mutat jelentős különbségeket a vizsgált időszakban.

Kísérleti személyek, anyag, módszer

Öt magyar anyanyelvű női beszélő (életkoruk átlaga 23 év) vett részt a kísérletben. Valamennyien ép hallásúak, beszédhibájuk nincsen, budapestiek, bölcsészhallgatók. A vizsgálat anyagát két középső és két alsó nyelvvállású, veláris és palatális hang ([ɔ, o, ɛ, e]) képezte. Valamennyi magánhangzó a tesztmondatok első tartalmas szavának hangsúlyos szótagjában jelent meg (az elemzett magánhangzókat félkövérítéssel jelöltük): *kérsz* „Kérsz egy falatot az almámból?”; *gondolod* „Azt gondolod, hogy Annának van igaza?”; *magyar* „A magyar úrturista kalandjait az egész országban figyelték.”; *gyermek* „A gyermekek bukfencezni is megtanulnak testnevelésórán.” A két rövidebb mondatban 23 és 28, a két hosszabban pedig 47 és 48 beszédhang realizálódott. Ez a négy mondat 15 jól formált mondatból álló anyag részét képezte.

Az adatközlők feladata az volt, hogy egyenként ismételjék meg a mondatokat, amelyeket a kísérletvezető (hasonló korú nő) felolvasott. Az ismétlések minden esetben egyszeri elhangzás után történtek. Az ismétléses módszert azért választottuk, hogy a spontán beszédhez szükséges felsőbb tervezési folyamatok ne legyenek aktívak a beszéd során, ezáltal az egyéni beszédtervezési különbségekből adódó, artikulációt érintő tényezők jelentős részét kiiktattuk. Az adatközlők hangos olvasási technikájának feltételezett eltérései miatt a felolvasástól is eltekintettünk. A mondatok hossza, a bennük előforduló szavak száma, valamint a tartalmak nagyjából azonos feldolgozási műveleteket biztosítottak. Az interjúkészítő kiejtésének tempója azonban hatással lehetett az adatközlők ismétléseire. Ezért statisztikailag elemeztük valamennyi mondatának a tempóját mind az egyes beszélők, mint pedig az alkalmak szempontjából. A különbségek igen kicsik (a tempóértékek többségükben teljesen azonosak); statisztikailag nem volt kimutatható közöttük szignifikáns eltérés.

A kutatás 7 hónapos követéses vizsgálat keretében valósult meg, a felvételeket kétheti rendszerességgel rögzítettük (összesen 14 alkalommal), mindig azonos helyen, stúdiókörülmények között, az ELTE Fonetikai Tanszékének csendesített helyiségében. A mondatismétléseket Zoom H4 felvevővel, beépített mikrofonnal vettük fel. Összesen 280 realizációt elemeztünk (5 beszélő \times 4 magánhangzó \times 14 felvétel, mondatidőtartam és tempó; összesen 840 adat).

A hanganyag feldolgozásához a Praat 5.3 szoftvert (Boersma–Weenink 2009) használtuk, ezzel végeztük el az annotálást, az adatolást és a méréseket. Elemeztük a mondatok időtartamát, a beszélők artikulációs tempóját (a mondatokat az adatközlők mondaton belüli szünet nélkül mondták vissza), valamint a négy kiválasztott magánhangzó időtartamát. A magánhangzókat folyamatos akusztikai visszacsatolás és vizuális ellenőrzés alapján azonosítottuk és szegmentáltuk. Az annotálást az egyik szerző végezte, míg a másik ellenőrizte azt. Amennyiben eltérés mutatkozott a két vélemény között, egy harmadik fonetikus véleményét kértük ki (a két vélemény közötti eltérés ke-

vesebb volt 3%-nál). A mondatok időtartamát az első hang kezdetétől az utolsó hang lecsengéséig, a magánhangzókét pedig a második formáns megjelenésétől ugyancsak annak lecsengéséig mértük. Az adatok automatikus ki-nyeréséhez egy erre a célra készült Praat-szkriptet alkalmaztunk, a kapott adatokat manuálisan ellenőriztük (és szükség esetén korrigáltuk). Az elemzések során a következő tényezőket vettük figyelembe: a magánhangzó minősége, a beszélő személy (ugyanazon beszélő és különböző beszélők), valamint a felvételi alkalmak.

A statisztikai elemzések során a General Linear Mixed Model eljárást alkalmaztuk. A függő változót a magánhangzók időadatai adták, a független változókat a magánhangzó minősége és a beszélők, valamint az alkalmak jelentették. Az alkalmakat statisztikailag három csoportba osztva vizsgáltuk (az első és az utolsó négy, valamint a közöttük lévő hat felvétel alapján). A mondatok ismétlésének temporális adatait statisztikailag a Friedman- és a Monte-Carlo-tesztekkel dolgoztuk fel (itt a függő változó a mondatok időtartama, illetve tempója, a független változók pedig a mondat, az alkalmak és a beszélő személyek voltak). Minden esetben az SPSS szoftver 19.0 változatát használtuk. Az elemzések 95%-os megbízhatósági szint tekintetbevételével készültek.

Eredmények

A mondatismétlés módszerével számos lehetséges változónak az időviszonyokra gyakorolt hatását csökkentettük. A kapott adatok temporális sajátosságai tehát döntően az idő múlásának tudhatók be.

A mondatok ismételt ejtésének temporális jellemzői

Az artikulációs tempó mérési adatait a másodpercenként ejtett beszédhangok számával fejeztük ki. Az 1. táblázat az öt női beszélő átlagos artikulációs tempójának értékeit, valamint az ismételt mondatok átlagos időtartamát és a szórásukat összegzi függetlenül a felvételi alkalmaktól.

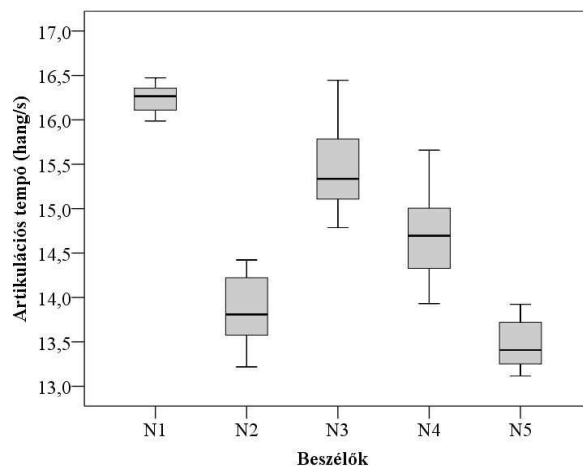
1. táblázat: Az összes mondat teljes időtartamának és az artikulációs tempónak az adatai beszélőnként

Beszélők	Mondat időtartama (s)		Artikulációs tempó (hang/s)
	Átlag	Szórás	Átlag
N1	2,22	0,50	15,46
N2	2,59	0,38	13,87
N3	2,30	0,27	13,48
N4	2,40	0,35	16,16
N5	2,63	0,49	14,65

A mondatok időtartama szignifikáns különbséget mutatott a beszélők és a mondatok függvényében (Friedmann- és Monte-Carlo-tesztek, $\chi^2(2) = 36,023$,

$p < 0,001$); az alkalmak nem mutattak ilyen hatást a mondatok időtartamaira. A mondatok teljes időtartamára nyilvánvaló hatással van a szavak, illetve a beszédhangok száma, így nem meglepő, hogy a négy mondat ejtési ideje szignifikánsan különbözik egymástól. (Az [e]-t, [ɔ]-t, [o]-t és [e:] -t tartalmazók átlagértékei: 3,08 s, 3,23 s, 1,79 s, 1,62 s; az átlagos eltérések: 0,23 s, 0,31 s, 0,26 és 0,12 s).

A tempóértékek is szignifikáns különbséget igazoltak az egyes beszélők között (Friedmann- és Monte-Carlo-tesztek, $\chi^2(2) = 70,0$, $p < 0,001$). (Mint-hogy az adatok nem normál eloszlásúak, ezért az egyes beszélők közötti esetleges különbségek mértékét statisztikailag nem tudtuk ellenőrizni.) A leglassabb és a leggyorsabb artikulációs tempójú mondat átlaga közötti különbség 2,68 beszédhang másodpercenként. Az N1-es és az N5-ös beszélő artikulációs tempója valamivel kevésbé változik a másik hároméhoz képest. Átfedéseket több beszélő esetén tapasztaltunk (1. ábra).

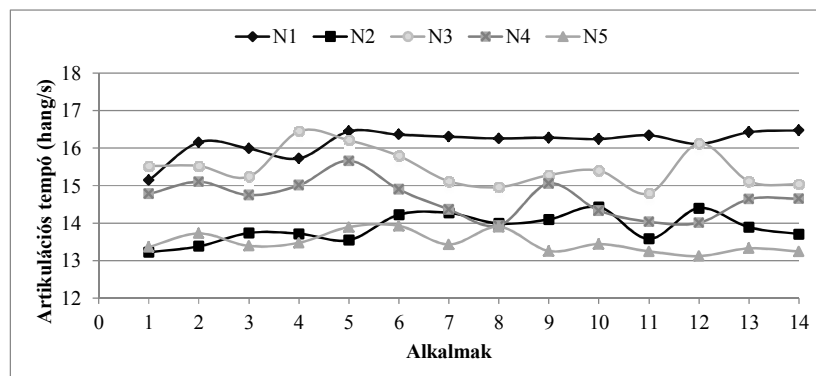


1. ábra

Az ismételt mondatok artikulációstempó-értékei (medián és szórás) az alkalmaktól függetlenül

Az eltelt idő függvényében sajátosan változnak az adatközlők artikulációs tempóinak értékei, noha a statisztikai elemzés az alkalmak tekintetében nem mutatott szignifikáns különbséget. A legkisebb tempóérték 13,11 hang/s, míg a legnagyobb érték 16,47 hang/s volt (2. ábra). A beszélők jellegzetesen különböztek abban, hogy az egymást követő felvételek között nagyobb vagy elenyésző tempóbeli eltéréseket produkáltak. Az N1, N2 és N5 jelű beszélők hosszabb időn át, néhány egymást követő felvételen alig mutattak változást a tempóértékeikben. Az N3-as és az N4-es beszélő artikulációs tempója ugyan-

akkor szinte semmilyen állandóságot nem mutatott. Mindennek sokféle oka lehet, a beszélő egyénisége éppúgy eredményezhette, mint egyéb (nem ismert) külső vagy belső tényezők. A mondatok ejtési időtartama és az artikulációs tempó nem mutatott statisztikailag releváns összefüggést.



2. ábra

Az artikulációs tempó értékeinek alakulása az egyes felvételi alkalmakkor

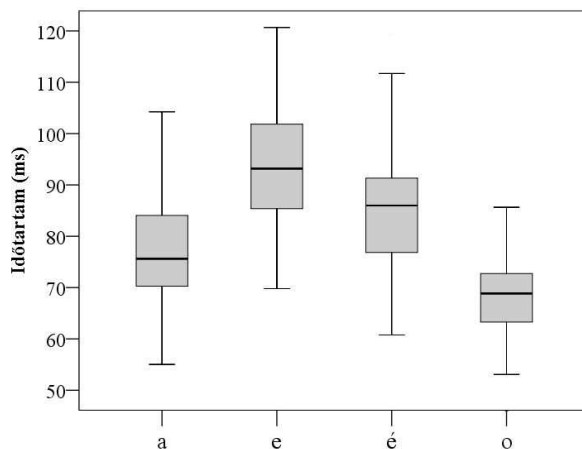
Tekintettel arra, hogy az alkalmak függvényében nem volt statisztikailag igazolható eltérés az artikulációs tempóban, ezért létrehoztunk egy skálát úgy, hogy az egyes beszélők produkciója során mért legkisebb artikulációs tempót tekintettük 100%-nak, majd ehhez viszonyítottuk a változásokat a különböző felvételi alkalmakkor. Az adatok szerint az N4-es beszélő esetében volt a legnagyobb arányú a változás az egyes felvételi alkalmak között, itt 12,42%-nak adódott a módosult érték. A legkisebb változást az N5-ös adatközlőnél találtuk, akinek az artikulációstempó-értékei csupán 6,1%-os változást igazoltak.

A magánhangzók időadatai

A következőkben először a beszélőktől függetlenül a magánhangzók időadatait elemezzük. A négy vizsgált magánhangzó közül egy fonológiaiag hosszú, három pedig fonológiai rövid hang, ezért – az átfedések mellett – azt vártuk, hogy a fonológiai különbség fonetikailag is tükröződik az átlagértékekben. Adataink azonban ezt a feltételezést nem teljesen igazolták. Az összes magánhangzó átlagos időtartama 82 ms-nak adódott (szórás: 15 ms); a legrövidebb magánhangzó 47 ms-os, míg a leghosszabb 136 ms-os volt. A fonológiaiag rövid magánhangzók átlagos időtartama 80,4 ms (szórás: 10,6 ms), a fonológiaiag hosszú magánhangzóé pedig 84,8 ms (szórás: 12,2 ms).

Az adatközlők a fonológiaiag rövid [ɛ] magánhangzót ejtették a leghosszabban (átlag: 95,8 ms, szórás: 13,5 ms), majd a fonológiaiag hosszú [e:] következett (átlag: 84,8 ms, szórás: 12,2 ms). Rövidebb volt a fonológiaiag

rövid [ɔ] (átlag: 77,6 ms, szórás: 11,0 ms), a legrövidebbnek pedig a fonológiai rövid [o] magánhangzó adódott (átlag: 67,9, szórás: 7,4 ms), vö. 3. ábra. Az adatolt hangidőtartamok szórása valamennyi magánhangzó esetében hasonló volt, a legkisebb érték az [o]-nál jelentkezett. A mintegy 13 évvel ezelőtti, hasonló vizsgálatban (Gósy 2002) ugyanezen magánhangzók fizikai időtartamai nem mutattak jelentős eltérést egymáshoz képest.



3. ábra

A négy elemzett magánhangzó időadatai (medián és szórás)

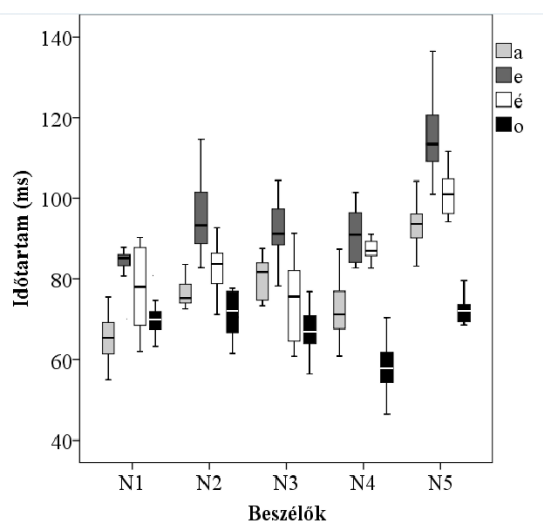
A statisztikai adatok mind a négy magánhangzó esetében igazolták, hogy a beszélő személye [$F(4, 252) = 64,416, p < 0,001; \eta^2 = 0,555$], valamint a magánhangzó minősége [$F(3, 252) = 150,966, p < 0,001; \eta^2 = 0,682$] szignifikáns hatással van az időtartamértékekre. A két tényezőnek az összefüggése is szignifikáns [$F(12, 252) = 8,397, p < 0,001; \eta^2 = 0,317$]. A parciális együtttható értékei arra utalnak, hogy a magánhangzó minőségének nagyobb a magyarázó hatása (közel 70%), mint a beszélő személyének (55%). Az egyes magánhangzók temporális adatait a 2. táblázat összegzi, valamint a 4. ábra szemlélteti. Az ábra az ugyanazon beszélő ejtésének és a különböző beszélőkéinek a különbségét demonstrálja. Noha relatíve nagyok az időtartamok realizációs tartományai az egyes beszélők magánhangzóinak esetében, ezek még mindig szűkebb tartományt jelölnek ki, mint a különböző beszélők értékei.

Noha az artikulációs tempó és a magánhangzók időtartama természetesen összefügg; ez az összefüggés számos tényező hatásának érvényesülése miatt bonyolult. A közepes artikulációs tempójú N5-ös beszélő ejti például a leghosszabb [ɛ, ɔ, ɛ:] magánhangzókat (a többi beszélőhöz viszonyítva), az [o]

magánhangzóinak időtartamai azonban nem térnek el jelentősen a többi beszélő adataitól.

2. táblázat: Az egyes beszélők magánhangzóinak temporális adatai

Magánhangzó	Időadatok (ms)	Beszélők				
		N1	N2	N3	N4	N5
[ɔ]	Átlag	65,3	76,9	80,3	72,1	93,5
	Határérték	55–76	73–84	73–88	61–87	83–106
[o]	Átlag	70,8	71,1	67,2	58,9	71,2
	Határérték	63–86	62–78	57–77	47–70	61–80
[e]	Átlag	84,3	95,6	91,5	91,9	115,9
	Határérték	70–96	83–115	78–104	83–95	101–136
[e:]	Átlag	77,9	82,9	74,8	86,4	101,8
	Határérték	62–90	71–93	61–91	76–95	94–120



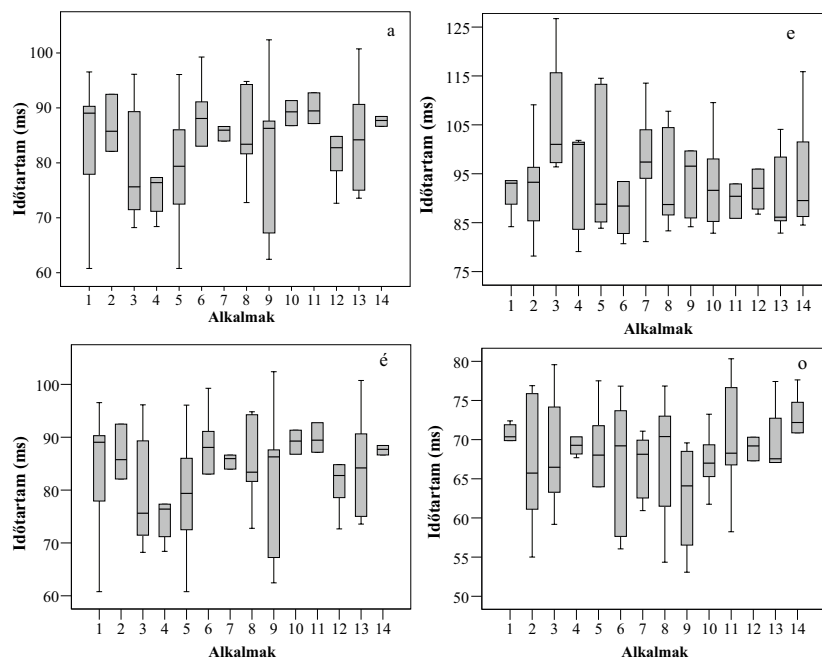
4. ábra

Az időtartamértékek (medián és szórás) változása a beszédhangminőség és a beszélő személyének függvényében

A csaknem azonos tempójú N2-es és N3-as beszélők [ɔ, o, e:] hangjainak időtartamai erőteljesen különböznek. A leggyorsabb artikulációs tempójú beszélő (N4) ejtésében csak az [o] adódott a legrövidebbnek, az [e:] magánhangzói például relatíve hosszúak. Az N1 és N2 jelű beszélők – artikulációs

tempójuk nagyobb különbsége ellenére – hasonló időtartamokban valósították meg a magánhangzókat.

Az 5. ábra a beszélőktől függetlenül szemlélteti a magánhangzó-időtartamok alakulását a **felvételi alkalmak**nak megfelelően az egyes magánhangzók minőségének függvényében. Az [ɔ] és az [e:] időtartamai nagyobb különbségeket mutatnak az egyes felvételek között, mint az [ɛ] és az [o] magánhangzókéi.



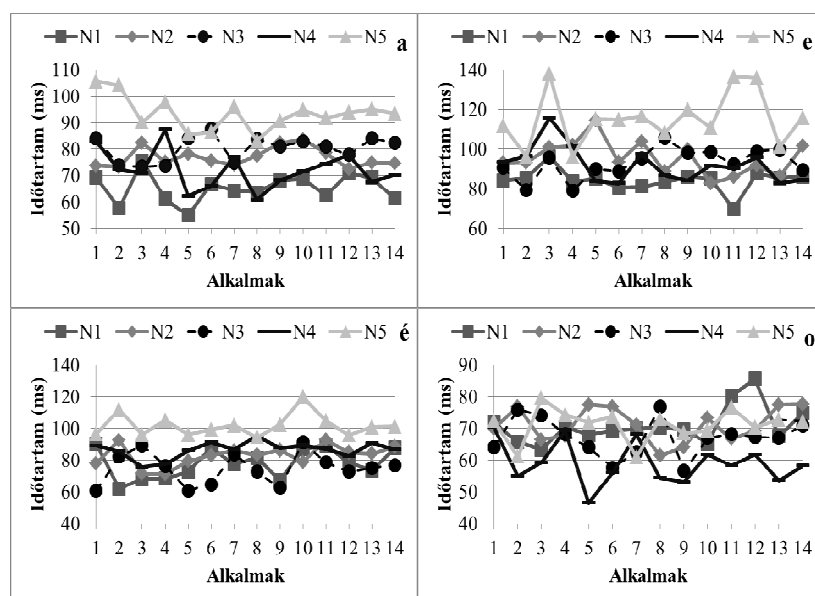
5. ábra

Az időtartamértékek (medián és szórás) a felvételi alkalmak és az egyes magánhangzó-minőségek szerint

Mint láttuk, a fonológiaiul hosszú [e:] magánhangzó időtartama nem realizálódott a leghosszabb fizikai időértékekkel. Az alkalmakat tekintetbe véve eltérőek a mért adatok mind beszélőn belül, mind beszélők között, azonban a különböző beszélők között nagyobbak az időtartam-különbségek, mint ugyanazon beszélők esetében. Ez egyértelműen mutatkozik, ha a felvételek egymásutánjában szemléltetjük az egyes beszélők és valamennyi elemzett magánhangzó időadatait (6. ábra).

A három felvételi időre osztott adatok nem mutattak szignifikáns különbséget az alkalmak tekintetében. Az első négy felvételen ejtett magánhangzók

átlag 81,9 ms (szórás: 15,4 ms), a következő hat alkalom alapján összegzett időadatok átlaga 80,7 ms (szórás: 15,3 ms), az utolsó négy felvétel adatainak átlaga pedig 82,4 ms-nak adódott (szórás: 14,9 ms).



6. ábra

A vizsgált magánhangzók időadatai a beszélők, a magánhangzók és az alkalmak tükrében

Következtetések

Kutatásunkban arra kerestünk választ, hogy vajon hét hónapos vizsgálati időtartamban mennyire változik négy magánhangzó időtartama és a hordozó mondatok artikulációs tempója mondatvisszamondásos kísérlet alapján. A kísérletben öt fiatal, női beszélő vett részt. Az eredmények azt igazolták, hogy csak a beszélő mint tényező, azaz a kiejtés, valamint a magánhangzó minősége van szignifikáns hatással a magánhangzók időtartamára, a felvételi alkalmak azonban nem. Ez azt jelenti, hogy a mért adatok olykor szembetűnő különbségei ellenére matematikailag nem igazolható eltérés az eltelt idő függvényében. A hét hónapi időmúlás, illetve a felvételek között eltelt két hét nem jelentett az artikulációra fordított időtartam szempontjából jelentős különbséget. Ez arra utal, hogy a beszélők mondatejtésének temporális szerveződése bizonyos határok között történik, és ugyanazon beszélő ezeket a határokat (nyilván tudat alatt) nem lépi át.

A magánhangzó minősége statisztikailag igazolhatóan gyakorolt hatást az időtartamok alakulására. Ugyanazon beszélők között kisebb variabilitást tapasztaltunk az időviszonyokban, mint a különböző beszélők között az elemzett magánhangzók időtartamában. Ez a megállapítás igaz a mondatok artikulációs tempójára vonatkozóan is. Feltételeztük, hogy a vizsgált időtartamban a magánhangzók fizikai időértékeire az állandóság jellemzőbb lesz, mint a változás. Ez a hipotézisünk egyértelműen teljesült. Levonható tehát az a következtetés, hogy a hét hónap alatt, mondatok ismétlése során ejtett [ɛ, ɔ, o, e:] magánhangzók időtartama stabil. Árnyalja a megállapításunkat az a tény, hogy a kéthetenként történt felvételek következtében óhatatlanul működtek bizonyos tanulási folyamatok, amelyek a stabilizálódás irányába hathattak. Az azonban az adatok alapján nem állítható, hogy a hetedik hónaphoz közeledve a magánhangzók időtartama kevésbé változatos lett volna (vö. 5. ábra).

Nem várt eredmény volt ugyanakkor, hogy a fonológiaiul hosszú [e:] magánhangzó fizikai időtartamai rövidebbek voltak, mint a fonológiaiul rövid [ɛ] magánhangzóéi valamennyi beszélő ejtésében. Az [e:] átlagos időértéke az egyes beszélőknél különböző mértékben tért el az [ɛ] időértékeitől, a legkisebb különbség 5 ms, a legnagyobb pedig 17 ms volt. Fiatal beszélők spontán beszédében végzett időtartammérésekben a fonológiaiul hosszú magánhangzó fizikai időtartamainak átlaga hosszabb volt a rövidekéinél (Bóna 2014). A jelen vizsgálat adatai alapján egyértelmű magyarázatot nem tudunk adni; feltehető esetleg, hogy a prozódiai jellemzők (hangsúly, kérdő hanglejtés), avagy a hordozó szó hossza eredményezték a rövidülést. Az [ɔ] csak egyetlen beszélő esetében mutatott hosszabb átlagidőtartamot, mint az [e:].

A második feltételezésünk (amely szerint az időbeli állandóságot a magánhangzó minősége befolyásolja), valamint a harmadik (hogy az ismételt mondatok artikulációs tempója nem mutat jelentős különbségeket), egyaránt teljesült. Annak ellenére, hogy relatíve nagyok voltak ugyanazon beszélők időtartamainak szórásértékai, minden esetben szűkebbnek adódtak, mint amikor az összes beszélő adatait vettük figyelembe. Várhatóan az [ɛ] magánhangzó fizikai időtartamértékei mutatták a legnagyobb beszélők közötti szórásértékot (66 ms), míg ugyanazon beszélők esetében a legnagyobb tartomány 35 ms volt, a legkisebb pedig 12 ms. A második legnagyobb szórásértékot az [e:] magánhangzónál tapasztaltuk (59 ms), de ugyanazon beszélők között itt is lényegesen szűkebb volt ez az értéktartomány, a maximális 26 ms-nak, a minimális 19 ms-nak adódott. Az [ɔ] esetében a beszélők közötti szórásérték 51 ms volt, ugyanazon beszélők esetében pedig 26 ms a legnagyobb és 11 ms a legkisebb értéktartomány. A különböző beszélők között is a legszűkebb szórást az [o]-nál mértük (39 ms), ugyanazon beszélők adatait tekintve a legnagyobb tartomány 23 ms, a legszűkebb pedig 16 ms volt.

Megállapítható tehát, hogy az időtartamok szórásában meghatározó a magánhangzó minősége, de egyértelműen nagyobbak a különbségek a különböző beszélők között, mint ugyanazon beszélő ejtéseit tekintve. A beszélők nem

konzekvensek abban, hogy ugyanazon magánhangzókat hasonló időtartamokban realizálják. A négy elemzett magánhangzó egyikének sincs ilyen kitüntetett ejtési jellemzője. A különböző beszélők magánhangzóinak időtartamszórásait összegezve és összevetve, a két elől képzett és a két hátul képzett között kismértékű különbség látható (mindössze 5-6 ms), az előbbiek a variábilisabbak.

Eredményeink lehetőséget nyújtanak arra, hogy jobban megismerjük a vizsgált magánhangzók időtartamainak variabilitását az eltelt idő függvényében, illetve ennek hatását az időzítési folyamatokra. Mindez a bűnügyi fonetikában hozzájárul a beszélő felismerésének fonetikai alapú megoldásaihoz.

Irodalom

- Beckford Wassink, Alicia – Wright, Richard A. – Franklin, Amber D. 2007. Intraspeaker variability in vowel production: An investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers. *Journal of Phonetics* 35. 363–379.
- Benzeguiba, M. – De Mori, R. – Deroo, O. Dupon, S. – Erbes, T. – Wellekens, C. et al. 2007. Automatic speech recognition and speech variability: A review. *Speech Communication* 49. 763–786.
- Boersma, Paul – Weenink, David 2009. Praat: doing phonetics by computer. <http://www.praat.org/> (A letöltés ideje: 2010. március 6.)
- Bóna Judit 2014. *A spontán beszéd sajátosságai az időskorban*. Eötvös Kiadó, Budapest.
- Drager, Katie 2010. Sociophonetic variation in speech perception. *Language and Linguistics Compass* 4. 473–480.
- Dromey, Christopher – Sanders, Marybeth 2009. Intra-speaker variability in palatometric measures of consonant articulation. *Journal of Communication Disorders* 42. 397–407.
- Gayraud, Frederique – Lee, Hye-Ran – Barkat-Defradas, Melissa 2011. Syntactic and lexical context of pauses and hesitations in the discourse of Alzheimer patients and healthy elderly subjects. *Journal of Clinical Linguistics and Phonetics* 25. 198–209.
- Gósy, Mária 2002. Long-term within-speaker and between-speaker differences in phonetic output: Evidence from Hungarian. In Braun, Angelika – Masthoff, Herbert R. (eds.): *Phonetics and its Applications. Festschrift for Jens-Peter Köster on the Occasion of his 60th Birthday*. Steiner, Stuttgart. 75–85.
- Gósy Mária – Beke András 2010. Magánhangzó-időtartamok a spontán beszédben. *Magyar Nyelvőr* 134/2. 140–165.
- Greenberg, Steven – Carvey, Hannah – Hitchcock, Leah – Chang, Shuangyu 2003. Temporal properties of spontaneous speech – a syllable-centric perspective. *Journal of Phonetics* 31. 465–485.
- Harmegnies, Bernard – Landercy, Albert 1988. Intra-speaker variability of the long term speech spectrum. *Speech Communication* 7. 81–86.
- van den Heuvel, Henk – Cranen, Bert – Rietveld, Tony 1996. Speaker variability in the coarticulation of /a, i, u/. *Speech Communication* 18. 113–130.

- Kahn, Juliette – Audibert, Nicolas – Bonastre, Jean-François – Rossato, Solange 2011. Inter- and intra-speaker variability in French: An analysis of oral vowels and its implication for automatic speaker verification. In: *Proceedings of the XVIIth International Congress of Phonetic Sciences*. University of Hong Kong, Hong Kong. 1002–1005.
- Kovács Magdolna 2002. *Tendenciák és szabályszerűségek a magánhangzó-időtartamok produkciójában és percepciójában*. Debreceni Egyetem Kossuth Egyetemi Kiadója, Debrecen.
- Lindblom, Björn 1990. Explaining phonetic variation: A sketch of the H&H theory. In Hardcastle, William J. – Marchal, Alain (eds.): *Speech production and speech modeling*. Kluwer, Dordrecht. 403–440.
- McDougall, Kirsty 2006. Dynamic features of speech and characterization of speakers: Towards a new approach using formant frequencies. *Speech Language and the Law* 13. 89–126.
- Mooshammer, Christine – Perrier, Pascal – Fuchs, Susanne 2008. Speaker-specific patterns of token-to-token variability. *The Journal of the Acoustical Society of America* 123. 3076.
- Recasens, Daniel – Espinosa, Aina 2006. Dispersion and variability of Catalan vowels. *Speech Communication* 48. 645–666.
- van Santen, Jan P. H. 1992. Contextual effects on vowel duration. *Speech Communication* 11. 513–546.
- Schötz, Susanne – Frid, Johan – Löfqvist, Anders 2012. A comparative study of kinematic and acoustic age-related variability in speech. In: *Proceedings of Fonetik 2012*. Gothenburg University. <http://lup.lub.lu.se/luur/download?func=downloadFile&recordId=3350408&fileId=3350431> (A letöltés ideje: 2014. augusztus 8.)
- da Silva, Paula Torres – Master, Suely – Andreoni, Solange – Pontes, Paulo – Ramos, Luiz R. 2011. Acoustic and long-term average spectrum measures to detect vocal aging in women. *Journal of Voice* 25. 411–419.
- Stevens, Kenneth 1999. *Acoustic phonetics*. MIT, Cambridge, MA.
- Zhang, Cuiling – van de Weijer, Joost – Cui, Jingxu 2006. Intra- and inter-speaker variations of formant pattern for lateral syllables in Standard Chinese. *Forensic Science International* 158. 117–124.

THE LARYNGEAL PROPERTIES OF SLOVAK THREE-CONSONANT CLUSTERS

Zsuzsanna Bárkányi – Zoltán G. Kiss

Introduction

In this paper,¹ we study the phonetic properties of three-consonant clusters (CC#C) in Slovak. More precisely, we will investigate the laryngeal properties of the velar–alveolar stop clusters /kt/ and /gd/, and the alveolar fricative–stop clusters /st/ and /zd/ in word-final position when followed by a voiced or a voiceless obstruent, or a sonorant consonant. This topic is of interest for two reasons: (i) there are not many studies dealing with the laryngeal characteristics of three-consonant clusters, and (ii) the study of consonant clusters can shed further light on the issue whether or not regressive voicing assimilation (RVA) in general, including pre-sonorant voicing, is a neutralizing process in Slovak.

We will seek to answer the following research questions: (i) Is voicing assimilation in Slovak (with obstruent and sonorant consonants as triggers) neutralizing or incomplete? (ii) Does pre-sonorant voicing in Slovak differ from pre-obstruent voicing? (iii) Is the word-final devoicing of obstruent clusters a completely neutralizing process in Slovak, or is it an example of incomplete laryngeal neutralization?

Background

Neutralization

Understood in its most well-known sense, phonological neutralization refers to the case when two or more contrastive sound segments suspend their contrast under specific conditions, whereby only a limited set of the contrastive segments can occur in a particular position. Examples include vowel reduction in English (where only certain vowels may appear in an unstressed syllable) and laryngeal neutralization (see Silverman 2012 for a detailed discussion of neutralization, as well as Jansen 2004). Neutralization processes are considered to be *complete* when there is no difference between the underlyingly contrasting members either in production or in perception for any of the possible phonetic correlates of a given contrast in a given context. That is, forms (e.g., voiced and voiceless obstruents) that are distinguishable in certain contexts (e.g., in intervocalic position) are phonetically completely

¹We would like to thank the two anonymous reviewers for their valuable suggestions.

indistinguishable in the neutralizing contexts (e.g., word-finally or in pre-obstruent position). Thus, in such positions a devoiced segment cannot be distinguished from an underlyingly voiceless segment either in its phonetic properties or in its phonological behaviour in any way.

Neutralization interpreted this way, that is, true phonetic neutralization, rarely results in homophones though. Firstly, it has been observed (e.g. Charles-Luce 1993; Kaplan 2011) that voicing alternation is more likely to be nearly neutralized – as opposed to completely neutralized – in contexts that would otherwise be semantically ambiguous. That is, phonological patterns are sensitive not only to contrasts among segments, but also to contrasts among individual lexical items.

Secondly, even if a neutralization process does derive homophony, it will rarely be the case that there is semantic ambiguity because languages resort to other strategies – especially in diachronic terms – to avoid homophony. Silverman (2012) discusses Korean, a language that has numerous neutralizing alternations but where the amount of homophony resulting from these alternations is surprisingly low. Korean counterbalanced the attrition of root-final consonantal values by resorting to root compounding. We can still assume though that processes such as voicing assimilation can be completely neutralizing phonetically.

Theoretically, there are at least three facets of phonetic neutralization (which themselves could be classified even further, see Dinnsen 1985, for instance). First, in the case of *complete neutralization*, there is no difference between the underlyingly contrasting members either in production or in perception for any of the possible phonetic correlates of the given contrast. This is the traditional generative assumption of neutralization, which – with the advent of more and more experimental work – turns out to be rarely the case.

Another possibility is that there is some systematic acoustic and/or articulatory difference between the segments in question, but this difference is *not perceived*, or at least speakers are not aware of the contrast. Allophonic differences and the first stages of sound change typically belong to this group (Dinnsen 1985). A subclass of this group is when there is a slight articulatory difference which does not manifest itself acoustically. Beňuš and Gafos (2007), using a combination of magnetometry and ultrasound, found that Hungarian transparent vowels that trigger back harmony (*hid* ‘bridge’) showed a more retracted tongue body posture than phonemically identical vowels that trigger front harmony (*víz* ‘water’) even in isolation (that is, not in a suffixed form, which would be a simple coarticulatory phenomenon). Note that no acoustic or perception study so far has shown any differences between them.

A further possibility is that the members of a “neutralized” contrast are not identical after all. Some production-acoustic features might remain that are consistently and significantly different in the contrasting sounds, and which

are perceived by speakers. Processes belonging to this group can be quite varied again: contrast preservation despite the loss of a primary acoustic cue might be fairly robust in some cases, while very weak in others (see Steriade's p-map theory, Steriade 2008).

The notions *neutralization*, *categoricity* and *graduality* are closely connected. A variation is generally thought to be *categorical* if it can be described with the categorical values of phonological features, i.e., when an alternation occurs between two discrete categories (e.g., voiced and voiceless) with no intermediate values. An alternation is thought to be *gradient* if the acoustic characteristics of the variants reflect values in between these categories (e.g., partly voiced), even if these in-between categories are systematic. Partial neutralization is gradient according to this view since some acoustic characteristics might signal more voicing (e.g. vowel length), while others might signal less voicing (e.g. phonation itself), for instance.

The phonological context regarding voicing neutralization studied in this paper is the word-final position. It has been reported for German (Port et al. 1981; O'Dell–Port 1983; Charles-Luce 1985), Catalan (Dinnsen–Charles-Luce 1984; Charles-Luce 1993), and Polish (Ślowiaček–Dinnsen 1985; Ślowiaček–Szymanska 1989) that word-final laryngeal neutralization leaves some residual cues to the phonological voicing of obstruents. However, Foutakis and Iverson (1984) and Kahlen-Halstenbach (1990) found that word-final devoicing is phonetically complete in German. Jassem and Richter (1989) report the same for Polish. Experimental evidence concerning voicing assimilation is varied. There is experimental work demonstrating that regressive voicing assimilation is non-neutralizing, and therefore it is a low-level, phonetic process (e.g., Charles-Luce 1993 on Catalan and Burton–Robblee 1997 on Russian). In contrast, Hallé and Adda-Decker (2011) found that whenever it occurs, voicing assimilation is categorical in French. Strycharczuk and Simon (2013) claim the same about West-Flemish. The issue of complete vs. incomplete laryngeal neutralization is far from being settled either empirically or theoretically. And there are very few experimentally-based studies that deal with pre-sonorant voicing.

Pre-sonorant voicing

Pre-sonorant voicing is a type of regressive voicing assimilation whereby a word-final voiceless/devoiced obstruent is assimilated in voicing to a following sonorant consonant or vowel in the next word. This process has raised recurrent interest among phonologists mostly due to the fact that the apparent trigger of voicing assimilation is a segment which is not contrastively specified for voicing. Phonetically considered, sonorants may be suitable triggers of regressive voicing assimilation as they are phonetically voiced and rather resistant to devoicing. Yet, typologically, pre-sonorant voicing is much less frequent than pre-obstruent voicing. In phonetically-based models this is explained by the passive or modal phonation of sonorants as opposed to the

active voicing of voiced obstruents (see especially Jansen 2004 and the references therein).

There are some interesting restrictions that seem to apply to pre-sonorant voicing, which do not apply to “regular”, pre-obstruent voicing assimilation: it typically occurs in languages which display final devoicing (this, however, does not mean that in all languages with word-final devoicing we will find pre-sonorant voicing as well). Pre-sonorant voicing is also generally restricted to the word-final (or syllable-final) position. Slovak is a language displaying both pre-obstruent and pre-sonorant voicing assimilation, as reported by Pauliny (1979) and Rubach (1994). In Slovak, a word-final obstruent is realized voiced if it is followed by a voiced obstruent (1c), or by a sonorant consonant or a vowel in the next word (1a). The latter process also applies to clusters, but is not operative within the word (1b).

(1) Voicing assimilation and pre-sonorant voicing in Slovak (Pauliny 1979: 152–153)

a. pre-sonorant voicing across word-boundary

vták letí [fta:g lɛci:] ‘bird is flying’

chlap ani nejedol [xlab aɲi ɲɛjɛdɔl] ‘man didn’t even eat’

ješ’ a piť [jɛʒ’ a pʲɪ] ‘eat and drink’

b. no pre-sonorant voicing within the word

tma [tma] ‘darkness’

kladivo [klaʝivo] ‘hammer’

astma [astma] ‘asthma’

chlap-mi [xlapmi] ‘man-INSTR’

c. regressive VA among obstruents

kto [kto] ‘who’, *gde* [gɟɛ] ‘where’

chlap dochodí [xlab doxoʝi:] ‘man comes’

hrad pri [ɦrat pri] ‘castle next to’

d. final devoicing

plod [plot] ‘fruit’

plot [plot] ‘fence’

As far as the trigger of pre-sonorant voicing is concerned, significant variation is observed among languages. In some languages – like Slovak, shown in (1), Kraków Polish (Rubach 1996) or West-Flemish (Strycharczuk–Simon 2013) – sonorant consonants and vowels pattern together and induce voicing assimilation. West-Flemish differs from the other Southern Dutch dialects in that in those dialects, as reported by de Schutter and Taeldeman (1986), only vowels voice the final fricative of the preceding word, while in West-Flemish, fricatives are voiced before sonorant consonants as well across

word-boundaries: *zes jaar* [zez ja:r] ‘six years’. Similarly to the Southern Dutch dialects, /s/-voicing in Ecuadorian Spanish is also induced only by vowels. Standard Peninsular Spanish is exactly the other way round: /s/ is voiced when followed by a voiced obstruent or a sonorant consonant. The process is not limited to word-final position, syllable-final /s/ also undergoes voicing (Hualde 2005).

It has been reported in a number of studies that pre-sonorant voicing targets only subclasses of obstruents: in Dutch only fricatives undergo voicing assimilation induced by sonorants (Simon 2010), in Spanish only /s/. An illustrative example is provided by Jiménez-Lloret (2008), who report a dialect continuum in Catalan: in Central Valencian there is no voicing of word-final consonants before vowels, Alguerés and the Valencian dialect of la Costera have sibilant voicing, in the Valencian dialect of Palmera – apart from word-final sibilants – alveolar affricates also become voiced in pre-vocalic position, Central Catalan has variable /f/ voicing as well, while in Alicantino all word-final obstruents undergo voicing when followed by a vowel.

Three-consonant clusters

As we have mentioned in the introduction, there are not many studies dealing with the laryngeal properties of three-consonant clusters. Here we briefly cite a few studies that discuss the focus of our investigation, namely voicing assimilation in three-member clusters.

Central Catalan shows an intriguing asymmetrical system: pre-vocalic voicing affects word- and prefix-final sibilants and stop + sibilant clusters to the exclusion of singleton stops (Bonet-Lloret 1998; Wheeler 2005; Strycharczuk 2012). As far as sonorant consonants are concerned, they affect all obstruents equally. This type of “undergoer asymmetry” is problematic for any phonetically and/or functionally-based explanation. There are several competing hypotheses as to why fricative voicing may be preferred over stop voicing before sonorants but none of them can straightforwardly account for the question why stop + fricative clusters undergo voicing while singleton stops do not. Note that any output-oriented rule or constraint-based formal analysis can easily account for this pattern. (A vowel will voice the word-final sibilant, which then will voice the preceding stop.) However, the general issue of pre-sonorant voicing remains a problem for these models, too.

Strycharczuk (2012) analyzes sibilant voicing using a diachronic phonetic-functional model. According to her, the process originated as intervocalic sibilant voicing rather than being a pre-vocalic voicing process. She claims that the pattern started off as intervocalic voicing that targeted delaryngealized sibilants. These neutralized final obstruents are less likely to resist voicing spill over from the neighbouring sounds, as no active devoicing gesture is executed to counteract voicing. Passive voicing may also be less perceivable in stops than in sibilants. The next diachronic step was that listeners re-interpreted intervocalic sibilant voicing as pre-vocalic. The final stage in the

development of the Catalan pre-vocalic voicing involved rule telescoping (Hyman 1975), when a voiced pre-vocalic sibilant becomes an input to VA, which operates independently in the language. In the case of Catalan not only the undergoer asymmetry is puzzling, but the trigger asymmetry as well: vowels only voice sibilants and sibilant-final clusters (like [ps], [ks]) while sonorant consonants cause gradient voicing in all obstruents with significant inter- and intra-speaker variation. If we assume that the right-hand environment does have an effect on the duration of passive voicing, but what is essential is the presence or absence of a voicing target (i.e., a delaryngealized final obstruent), we do not expect any differences between sonorant consonants and vowels as triggers of pre-sonorant voicing. Strycharczuk (2012) hypothesizes that pre-vocalic voicing is the older pattern of the two in Catalan, which is supported by the high amount of variation observed in the pre-sonorant consonant process as well as the assumption that vowels are more conducive of passive voicing as they are more open.

Recasens and Mira (2013) also examine Catalan from an articulatory perspective, but they focus on $C_1C_2\#C_3$ sequences where C_2 is always an obstruent, while C_1 and C_3 may be an obstruent or a sonorant. The goal of their study is to investigate the extent to which word-final obstruents assimilate in voicing to the following word initial voiced consonant. The authors work within the Degree of Articulatory Constraint (DAC) model of coarticulation, which is based on the principle that the extent to which consonants resist the coarticulatory effects of other phonetic segments (*coarticulation resistance*) and exert coarticulatory effects on these adjacent segments (*coarticulation aggressiveness*) ought to increase with the involvement of a given articulator in their production. Thus, for example, since the tongue dorsum is more actively involved in the production of palatal consonants than in the case of labials and alveolars, the former consonants ought to be more resistant to tongue dorsum coarticulation effects from the adjacent vowels than the latter, while at the same time exerting more prominent coarticulatory effects on the vowels in question. Similarly, consonants which – because of their production requirements – are more prone to exhibit overall voicing are the ones that ought to be the most resistant to changes in voicing degree induced by the adjacent consonants and should also be the most aggressive as triggers of voicing. Thus, for example, sonorants (nasals, laterals) are expected to exert more voicing coarticulation on preceding obstruents (stops, fricatives) than obstruents since they exhibit more voicing and are less prone to devoice across contextual conditions.

In an earlier study (Recasens–Mira 2012), the authors found, contrary to the initial expectation, that syllable-final fricatives and stops showed much less voicing than expected before nasals and laterals (above 80% voicing in C_2 , less than 45% voicing in C_1), and voicing differences as a function of place of articulation did not extend into C_1 . Note that C_2 in this case is the

target consonant not the trigger. According to the authors, the presence of little voicing during obstruents followed by nasals and laterals appears to be due to the need to preserve the pressure difference across the oral constriction for intense turbulence and thus the integrity of the frication noise for fricatives, and to allow for a sufficient intraoral pressure build-up for the generation of a salient burst for stops, which could be impaired if regressive voicing occurred simultaneously with anticipatory nasalization for nasals and with anticipatory tongue front raising for laterals.

Data for three-consonant clusters reported by Recasens–Mira (2013) show lower percentages of vocal fold vibration in all three consonants as a general rule. Thus, voicing percentages across speakers and contextual conditions for syllable final obstruents subjected to voicing assimilation amounted to 5–45% in CCC sequences and to 30–45% in CC sequences in the case of fricatives, and to 5–55% in three-consonant clusters and to 55–60% in two-consonant clusters in the case of stops. These percentages confirm the hypothesis stemming from DAC that the degree of voicing should decrease with the number of consonants in the cluster and thus with an increase in the aerodynamic and articulatory demands involved. Consonant voicing percentages in three-consonant clusters differ considerably as a function of manner and place of articulation. Voicing coarticulation effects from specific consonants on others yielded little support for the Catalan regressive voicing rule, as the contribution of C_3 to voicing in the preceding syllable/word-final consonants was relatively small and did not always agree with the initial prediction that regressive voicing should increase with voicing degree in the triggering consonant. In particular, there was little voicing during obstruents when followed by a nasal or a lateral, which contradicts DAC, as the authors speculate, perhaps in order to allow for sufficient intra-oral pressure build-up for the generation of turbulent airflow and a burst which could be impaired by anticipatory nasalization for nasals and an earlier apical constriction for laterals.

Duration data reveal that the effect in question may be accompanied by C_2 shortening mostly when C_3 is a nasal. The patterns of voicing interaction between C_1 and C_2 lend some support to the hypothesis that voicing effects should be stronger if involving consonants located within the same syllable and word than across a syllable and word boundary. C_3 stop burst duration was also greater for clusters with a voiceless C_3 than for those with a voiced C_3 in stop + /s/ + stop clusters. Duration effects associated with the C_3 voicing distinction could not be traced during C_1 or the vowel preceding the cluster. These segment duration and intensity data suggest that speakers of languages where voiced stops exhibit voicing lead may use not only vocal fold vibration but other phonetic characteristics that depend more closely on air pressure and airflow for cueing the voicing contrast in clusters – as supported by a number of studies from different languages. The vocal fold vibration and segmental duration and intensity data just summarized indicate that, contrary

to current descriptive and phonological accounts, voicing assimilation in Catalan three-consonant clusters with a voiced C_3 cannot be modelled as a purely regressive process (Wheeler 2005). C_3 -dependent regressive voicing effects occur less than predicted by the phonological rule: obstruents are mostly voiceless when occurring in C_1 . C_3 position effects extend to some extent into C_2 but barely into C_1 . Vocal fold vibration data provide some support for voicing dependency between C_1 and C_2 and thus consonants placed in the same syllable final position. It thus appears that voicing assimilation may be conditioned by syllable and word affiliation as well. Moreover, considerable voicing effects between the two syllable final consonants occur at the progressive but not at the regressive level.

Markó et al. (2010) investigated CC and CCC clusters within the word and across the word-boundary in spontaneous and read speech in Hungarian. Here we only mention clusters that were not interrupted by pause of any length. Measurements were carried out manually by the authors, and realizations were classified into three groups: a consonant was considered as voiced if it contained a quasi-periodic signal in at least 80% of its duration. A consonant was considered as voiceless if it contained quasi-periodic signal in at most 20% of its duration. Between these values the consonant was designated to be partially voiced. It is somewhat difficult to evaluate the results of this study since both the manner and the place of articulation of the members of CCC clusters were quite varied in the spontaneous corpus, as well as the number of occurrences. The demonstrative *azt* 'that-ACC', for instance, was highly overrepresented, and there were very few tokens with three obstruents. There was much inter- and intra-speaker variation as well. Nonetheless, the authors conclude that for partially voiced realizations, a large difference is found between the voicing and devoicing types of assimilation. They assume that this difference is due to articulatory concomitances like the interaction of voicing assimilation and the physical constraint of devoicing, their reasoning is similar to Recasens–Mira (2013). They claim that the variability of the data confirms that Hungarian voicing assimilation is a gradient and sometimes only partly regressive process. They also observe that the process most of the time seems to operate obligatorily and that speech style can override it.

Singleton consonants in Slovak

In the remainder of this section, we will briefly summarize our earlier findings on the laryngeal properties of word-final alveolar obstruents in Slovak (Bárkányi–G. Kiss 2012, 2013). In Slovak, word-final single /t/, /d/, /s/ and /z/ were realized completely voiceless before a silent pause (with over 90% of unvoiced frames for all target consonants under scrutiny). There was no statistically significant difference between the voiced and the voiceless obstruents: /t/ vs. /d/: $b = 0.444$, $t(15) = 0.25$, $p = 0.806$; /s/ vs. /z/: $b = 2.692$, $t(15) = 1.54$, $p = 0.143$ (Figure 1). (The methodology of the experiment on

singleton consonants summarised here is identical to that presented in the section *Experiment* below.)

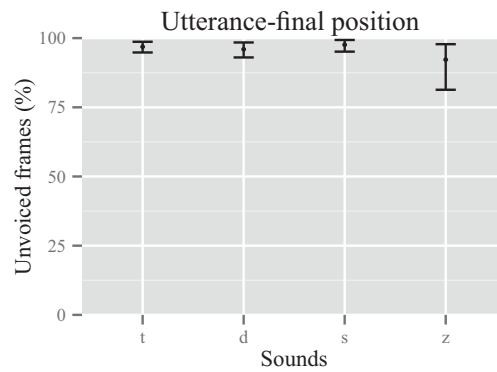


Figure 1.

Means of the ratio of the unvoiced part to total consonant length in Slovak utterance-final /t d s z/ (error bars indicate 95% confidence intervals)

We found a statistically significant difference in the case of /s/ vs. /z/ for two of the acoustic correlates of the voicing contrast. Their duration was not significantly different; however, the duration of the preceding vowel turned out to be significantly different [$b = -6.619$, $t(15) = -2.95$, $p = 0.0099$, effect size: $r = 0.61$], and consequently the vowel-to-consonant (V : C) duration ratio was also significantly larger for /z/ than for /s/ [$b = -0.051$, $t(15) = -2.44$, $p = 0.028$, effect size: $r = 0.53$]. Table 1 sums up the phonetic variables measured in utterance-final position and whether the members of each obstruent pair differed in a statistically significant way for them.

Table 1: Acoustic correlates of obstruent voicing in utterance-final position
* stands for a statistically significant difference ($p < 0.05$)

Acoustic correlates	/t/–/d/	/s/–/z/
Unvoiced frames		
Voicing duration		
Consonant duration		
Preceding vowel duration		*
V : C duration ration		*

Note that contrast-preservation between the voiced–voiceless pairs in word-medial intervocalic position is robust in the language. Word-final utterance-medial consonants, i.e., obstruents followed by another consonant in the next word showed sporadic significant differences between the acoustic

properties of the voiced–voiceless members of the alveolar obstruent pairs, but mostly point to the direction of voicing neutralization. Before /p/, the stops /t/ and /d/ differed with respect to the vowel duration variable only (not even in V : C ratio); before /b/ we found no difference in the case of stops, but /s/ and /z/ differed in V : C duration ratio.

In pre-sonorant position the fricatives were found to be statistically different in the amount of voicing, in consonant duration and vowel duration as well, although not in their ratio. Note that both /s/ and /z/ were realized with a fair amount of voicing (mean percentage of unvoiced frames for /s/: 20.95%, SD = 33.64%, for /z/: 39.32%, SD = 39.49%; mean voicing duration for /s/: 44 ms, SD = 18 ms, for /z/: 39 ms, SD = 22 ms); therefore, we might suspect that both fricatives are perceived as voiced by speakers, but this must be backed up by a follow-up perception experiment. Furthermore, both the vowel and the fricative itself were longer in the case of the underlyingly voiced fricative /z/, which partly contradicts universal trends. Figure 2 sums up the mean percentages of voicing in the final alveolar obstruents in Slovak in three assimilation environments.

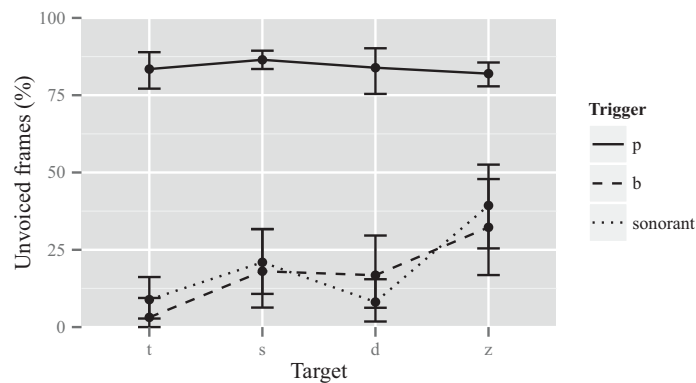


Figure 2.

Interaction graphs showing the mean ratio of the unvoiced part to total consonant length in word-final utterance-medial /t d s z/ followed by the voiceless obstruent /p/, the voiced obstruent /b/ and the sonorant consonants /m l/ in Slovak (error bars indicate 95% confidence intervals)

As mentioned above, the voicing distinction in Slovak single consonants seems to be almost fully neutralized in utterance-final position (final devoicing), as well as before voiced and voiceless consonants. Note that sonorants in Slovak have as much “voicing power” as voiced obstruents. Word-final obstruents do not differ in their voicing (unvoiced frames, voicing duration) whether they are followed by /b/ or sonorants, but they are significantly less

voiced when followed by /p/ [t/ followed by /b/ vs. the sonorants: $b = -2.90$, $t(10) = -1.13$, $p = 0.284$; /t/ followed by /b/ and the sonorants vs. /p/: $b = 25.80$, $t(10) = 16.0$, $p < 0.001$, effect size: $r = 0.98$; /s/ followed by /b/ vs. the sonorants: $b = -1.61$, $t(10) = -0.430$, $p = 0.675$; /s/ followed by /b/ and the sonorants vs. /p/: $b = 22.25$, $t(10) = 9.63$, $p < 0.001$, effect size: $r = 0.95$]. This indicates that sonorants in Slovak do not form an intermediate category as triggers of voicing assimilation. Pre-sonorant voicing in Slovak clearly and categorically patterns with pre-voiced obstruent voicing.

The present study aims to further investigate whether voicing assimilation in Slovak is really taken to the “end”, that is to say, whether it really is a stabilized categorical process, or it is more of a low level coarticulatory/phonetic phenomenon. We assume that if the process is categorical (including final devoicing), it should be neutralizing. If, on the other hand, voicing assimilation in Slovak is coarticulatory, the absolute duration of the voiced part across single consonants and consonant clusters should be fairly constant.

Experiment

Material

Words ending in /kt/–/gd/ and /st/–/zd/ were tested in the following three positions: (i) absolute word-final (utterance-final) position; (ii) word-final sentence-medial position, where the target obstruents were followed by one of the following triggers: (a) voiced obstruent /b/, (b) voiceless obstruent /p/, (c) sonorant consonant (/l/ or /m/), and (iii) sentence-medial intervocalic position.

The target obstruents were always preceded by the vowel /a/ or /o/; in intervocalic position, the vowel following the target consonants was /a/. The test words were *kontrakt* ‘contract’, *smaragd* ‘emerald’, *chvost* ‘tail’ and *drozd* ‘blackbird’. The use of minimal pairs was avoided on purpose because in our experience, despite the use of a fair number of distractors, subjects tend to overemphasize the differences in their pronunciation. Stimuli were embedded in carrier sentences: e.g., *Kontrakt bez pečate je neplatný* ‘The contract without a stamp is invalid’. The carrier sentences were 10–13 syllables long, neutral sentences, the target and the trigger occurred in the same intonational phrase; word-stress in Slovak falls on the first syllable. We did not find a significant difference in the behaviour of /l/ and /m/ with regard to their voicing capabilities, therefore we decided to collapse the data from pre-/m/ and pre-/l/ positions together into a common ‘pre-sonorant consonant’ context.

Methods

Six native speakers of Slovak participated in the experiment aged 20–52, none of them reported any speaking, hearing or reading disorder. They were all naive as to the aims of the experiment and participated as a courtesy to the authors. Subjects read the test sentences and fillers from a monitor screen in a

randomized order, which was generated by SpeechRecorder.² Each test sentence was read five times, but the first reading was considered as the familiarization phase, and was not taken into consideration. We investigated six contexts for four words by six subjects with four repetitions, which resulted in altogether 576 test items. Recordings were made in a sound-attenuated room with a Sony ECM-MS907 microphone connected to a laptop through an M-Audio MobilePre USB preamplifier external sound card. The material was recorded at a 44,100 Hz sampling rate, and was resampled at 22,050 Hz for the various acoustic measurements.

Measurements

The acoustic analysis was carried out in Praat (version 5.3.12, Boersma–Weenink 2012), for the statistical analysis we used R (version 2.15.0). The spectrograms were segmented manually by the authors and the following measurements were carried out on the basis of the inserted boundaries:

- (2) Correlates of laryngeal contrast measured in the experiment
 - a. *Phonation*-related correlates of laryngeal contrast
 - i. the absolute length of the voiced interval
 - ii. ratio of the unvoiced part compared to the total length of the consonant (“% of unvoiced frames”)
 - b. *Duration*-related correlates of laryngeal contrast
 - i. duration of the preceding vowel
 - ii. duration of the target consonant
 - iii. vowel-to-consonant duration ratio

Voicing was measured manually, based on the visual inspection of the spectrograms and oscillograms. In the case of stops, voicing was measured during the closure phase, i.e., up to the burst, but the release phase was not included (similarly to the methodology applied by Strycharczuk 2012, for instance). In the case of fricatives, voicing was measured during the whole duration of the frication noise. We measured two parameters: the absolute length of the voiced interval within the target consonants in seconds (referred to as “voicing duration” in the figures below) and the ratio of the unvoiced part compared to the total length of the consonant (referred to as “unvoiced frames” in the figures below).

As for the duration-related parameters, in the case of fricatives, the interval of frication noise was measured. In the case of stops, closure duration and release burst duration were measured. Since absolute segment durations are highly variable due to different speaking rates, the vowel-to-consonant dura-

² <http://www.bas.uni-muenchen.de/Bas/software/speechrecorder/>

tion ratio was also measured. It has been observed (e.g., Port–Dalby 1982) that the ratio between vowel duration to stop closure or fricative constriction remains relatively constant in words with the same underlying voicing feature: the vowel-to-consonant duration ratio is generally larger for voiced obstruents than for voiceless obstruents. Many perception-driven accounts derive the inverse patterning of voiced–voiceless obstruent length and preceding vowel duration as a form of mutual auditory enhancement for the voicing contrast. The idea is that increased vowel duration makes the duration of a following obstruent appear shorter, and conversely that a decrease in vowel duration increases the perceived duration of a following obstruent, and that vowel duration and obstruent duration are therefore integrated into a single percept (Port–Dalby 1982; Massaro–Cohen 1983; Kluender et al. 1988; Port–Leary 2005). This hypothesis has been largely supported by experimental evidence. Thus, listeners pay attention especially to the relative duration of a vowel and the constriction duration of a following obstruent (Javkin 1976; Parker et al. 1986; Kingston–Diehl 1994), which may serve to preserve the voicing contrast in phonetically unfavourable positions, known as “pre-fortis clipping” in the English phonetics/phonological literature (e.g., Wells 1982). For an extensive overview on the choice of these (and other) acoustic correlates of voicing contrast and voicing assimilation, see Jansen (2004), Bárkányi–Kiss (2007) and Strycharczuk (2012), and the references therein. In the present paper we will not discuss other cues that are also cited in the literature as correlates of laryngeal contrast, such as intensity of the frication and the burst, or the f_0 and F_1 of the surrounding vowels.

Statistical analysis

The acoustic correlates of voicing were analyzed with linear mixed-effects models, using the *nlme* package in R (Pinheiro et al. 2013; for the method see Field et al. 2012). For each phonetic outcome variable a mixed-effects model was fitted with random intercepts for subjects. The fixed parts of the model were: target clusters and trigger sounds (when a following sound followed the target cluster). The contrast coding (using planned orthogonal contrasts) distinguished between 1. target class: stop + stop (/kt, gd/) vs. fricative + stop (/st, zd/); 2. /kt/ vs. /gd/, and 3. /st/ vs. /zd/. The effect size measure used in the paper is Pearson’s correlation coefficient r .

Miscellaneous issues

We discarded from the analysis those cases where a silent pause of any length appeared between the word-final cluster and the following consonant in the sentence-medial word-final position. This amounted to 6.4% of all the tokens; such pauses were observed in the speech of two subjects. Four instances had to be discarded due to a technical error, thus 548 items could be analysed eventually.

For the current experiment we aimed to investigate the same type of final consonants (alveolar stops and fricatives) as in our earlier study on singletons (see the previous section and Bárkányi–G. Kiss 2012; 2013) so that the two sets of results could be consistently compared across the two studies. Furthermore, our aim was to elicit a fairly natural speech tempo. These factors partly contributed to C_2 being deleted in over half of the cases. In stop + stop clusters before C_3 , C_2 was deleted in 66.07% of the cases. It was in only 16.98% that both C_1 and C_2 were kept and fully released. The percentage of C_2 deletion in clusters composed of a fricative and a stop was 66.41%, but C_2 was kept and fully released in 30.79% of the cases.

Results and discussion

Utterance-final position

In this context both consonants (C_1 and C_2) were preserved. (This was not the case in word-final utterance-medial context as mentioned above in the previous section). The duration of the clusters was on average 145–158% of that of singleton consonants. In this position we did not find statistically significant differences between the underlyingly voiced vs. voiceless clusters for any acoustic cues. [Stops, unvoiced frames: $b = 1.012$, $t(15) = 1.304$, $p = 0.211$; fricatives, unvoiced frames: $b = -0.335$, $t(15) = -0.436$, $p = 0.669$; stops, vowel duration: $b = -0.002$, $t(15) = -1.48$, $p = 0.159$; fricatives, vowel duration: $b = 0.0004$, $t(15) = 0.273$, $p = 0.788$; stops, consonant duration: $b = 2.325$, $t(15) = 0.69$, $p = 0.5$; fricatives, consonant duration: $b = 0.04$, $t(15) = 0.012$, $p = 0.99$]. As Figure 3 exhibits, all the clusters were over 95% de-voiced in utterance-final position.

Intervocalic position

The intervocalic word-medial position was included in the experiment for the following reason. We assumed that this context was a phonetically “favourable” position, where contrast-preservation should be relatively robust.³ Our results backed up this expectation. Figure 4 shows the differences in the voicing ratio of the clusters [/kt/ vs. /gd/: $b = 47.64$, $t(15) = 26.40$, $p < 0.001$, $r = 0.99$; /st/ vs. /zd/: $b = 38.99$, $t(15) = 21.61$, $p < 0.001$, $r = 0.98$].

In this position, similarly to the absolute final context, both members of the cluster were maintained, we found no deletions here. We observed that in 20% of the cases, /g/ was not realized as a stop but rather as an approximant without a closure phase and a noticeable release burst. Vowel duration again showed a similar pattern to singletons. It was before /kt/ and /gd/ that vowel length did not significantly differ [/kt/ vs. /gd/: $b = -2.65$, $t(15) = -0.89$, $p =$

³ On the phonetically favourable nature of the prevocalic, intervocalic context for contrast preservation, see, among others, Steriade (1997); Hayes (1999); Hayes–Steriade (2004); Wright (2001, 2004).

0.3849; /st/ vs. /zd/: $b = -8.77$, $t(15) = -2.97$, $p = 0.0096$, $r = 0.61$), but overall if we consider the vowel–consonant duration ratio, both the stop and the fricative cluster pairs showed shortening effects; see Figure 5 [/kt/ vs. /gd/: $b = -0.082$, $t(15) = -3.21$, $p = 0.0058$, $r = 0.64$; /st/ vs. /zd/: $b = -0.19$, $t(15) = -7.33$, $p < 0.001$, $r = 0.88$].

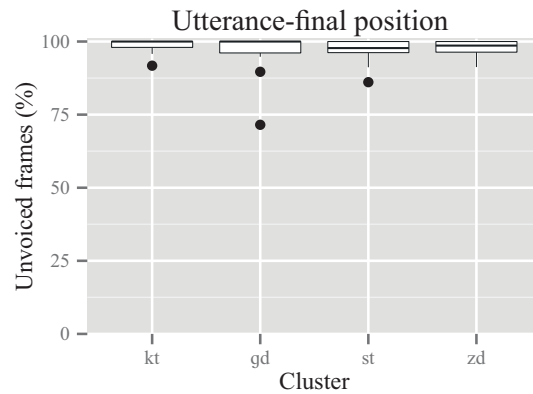


Figure 3.

Boxplots showing the ratio of the unvoiced part to total consonant length in Slovak utterance-final /kt gd st zd/ clusters

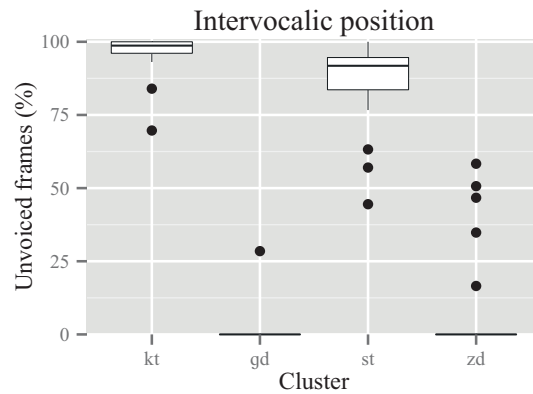


Figure 4.

Boxplots showing the ratio of the unvoiced part to total consonant length in Slovak word-medial intervocalic /kt gd st zd/ clusters

We conclude that while there seems to be complete voicing neutralization in utterance-final position, intervocalic obstruent clusters are fully contrastive

in the language. Let us now turn to the focus cases of this paper, namely, the voicing properties of CC#C clusters.

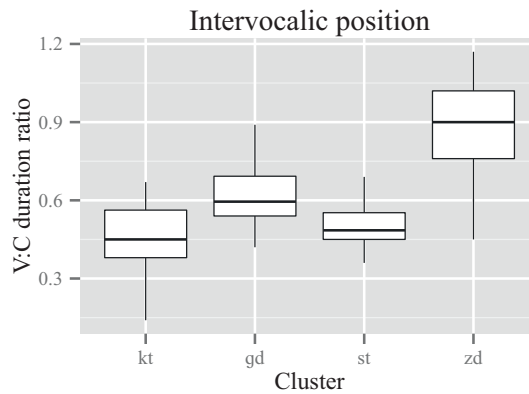


Figure 5.

Boxplots showing the V:C ratio in Slovak word-medial intervocalic /kt gd st zd/ clusters

CC#C clusters

Note that this is the context where in many cases we observed C_2 deletion. First, we will present the results for those cases where C_2 was preserved so that we can observe the voicing properties of consonant clusters in pre-consonant position. And then, we will proceed to compare voicing in C_1 in those cases where C_2 was preserved with those realizations where it was deleted.

We begin with those cases where we expected devoicing, that is, obstruent clusters followed by /p/ in the next word. Our expectations were borne out: all the clusters were realized with around 90% unvoiced frames, and we found no statistically significant differences between the underlyingly voiced vs. voiceless clusters [stop + stop clusters $b = -10.715$, $t(9) = -1.615$, $p = 0.14$; fricative + stop clusters $b = -0.755$, $t(9) = -0.138$, $p = 0.893$]. We did not find significant differences for any of the duration-related correlates either. The situation is very similar for the voicing correlates in the pre-/b/ and pre-sonorant contexts: comparing the voicing ratio of stop + stop clusters before /b/: $b = 19.128$, $t(4) = 1.859$, $p = 0.136$; and the fricative + stop clusters before /b/: $b = -12.78$, $t(4) = -1.414$, $p = 0.23$. Again, we did not find statistically significant differences for any of the duration-related parameters either. As for the pre-sonorant position, stop + stop clusters seem to be neutralized for voicing: $b = -1.163$, $t(11) = -0.239$, $p = 0.814$.

In the case of fricative clusters, we did find a significant difference; however, it occurred unexpectedly: it was /zd/, and not /st/, that displayed less voicing; actually, these clusters were much less voiced than expected (mean

unvoiced frames: 50%) compared to /st/ [average unvoiced frames: 24.19%]: $b = -11.282$, $t(11) = -2.4$, $p = 0.0348$, $r = 0.96$. We think that these unexpected results were a consequence of the small number of tokens we could measure (due to the deletion of C_2), and therefore, it is difficult to draw meaningful conclusions for these findings. We sporadically obtained statistically significant results for other variables, too, for similar reasons (low number of tokens). While we did not find significant differences either in vowel length or in consonant length, their ratio turned out to differ significantly for stop + stop clusters in pre-sonorant position, although the effect size of this significance was very low: $b = 0.123$, $t(11) = 2.436$, $p = 0.033$, $r = 0.038$. We note again that for the underlyingly voiced cluster /gd/ we found a lower value (duration ratio = 0.807) than for /kt/ (duration ratio = 1.00), which contradicts the usually observed tendencies for vowel-to-consonant duration ratio (for comparison, see Figure 5, which illustrates the findings of this duration variable in intervocalic position).

A much more interesting question is whether we can find differences in the voicing “aggressiveness” of /b/ vs. sonorant consonants, that is, whether the clusters under scrutiny are more voiced before an actively voiced obstruent than before a modally voiced sonorant. Another question related to the categoricity of voicing assimilation and within that pre-sonorant voicing in Slovak concerns whether the voicing properties of singleton consonants differ from those of consonant clusters. If we find important differences, we might assume that voicing assimilation is coarticulatory after all. If, however, the voicing properties of C vs. CC targets are similar, it points to the direction of a categorical process. Note that in this latter case there still might be a phonetic, aerodynamic difficulty in implementing voicing for a longer time, which can give rise to some differences. Let us compare the results on the voicing behaviour of singleton consonants (our earlier study cited above) and consonant cluster targets (present experiment).

We divided our data according to the following parameters: (i) stops vs. fricatives; (ii) singletons vs. clusters; (iii) clusters with deletion vs. no deletion. In this way we obtained the following “target class” groups: single stops (labelled as “singST” in the graphs below), single fricatives (“singFR”), stop + stop clusters with no deletion (“NoDelST”), stop + stop clusters where C_2 is deleted and therefore they are realized as single stops (“DelST”), fricative + stop clusters with no deletion (“NoDelFR”), and fricative + stop clusters where C_2 is deleted and therefore they are realized as single fricatives (“DelFR”). Figure 6 summarizes the amount of voicing in the six target classes in four different contexts (trigger classes): before sonorant consonants, before /b/, before /p/ and “nothing”, which stands for the absolute final position where there is no triggering segment. Note that in this final context, both members of the clusters were systematically articulated and therefore the groups “DelST” and “DelFR” are not applicable here.

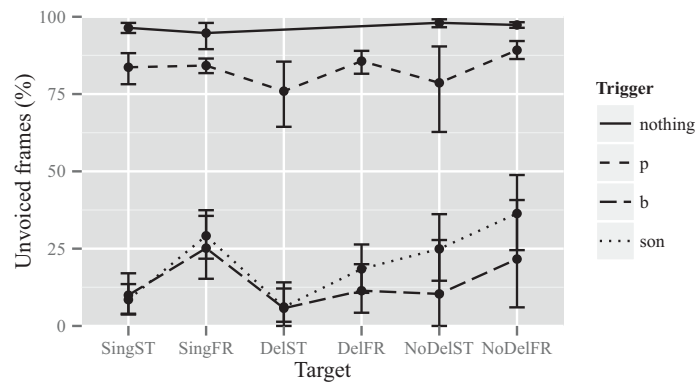


Figure 6.

Interaction graphs showing the mean proportion of voicing in word-final utterance-medial /t d s z/ and /kt gd st zd/ followed by the voiceless obstruent /p/, the voiced obstruent /b/ and the sonorant consonants /m l/, as well as in utterance-final position in Slovak (error bars indicate 95% confidence intervals)

As far as the voicing aggressiveness of /b/ vs. sonorant consonants is concerned, if we add all six target groups up, we do find a statistically significant difference with a medium effect size [$b = -3.21$, $t(26) = -2.231$, $p = 0.0345$, $r = 0.53$], despite the fact that both trigger full voicing (mean 19% of unvoiced frames in the case of sonorants and 14% for /b/). However, there are no interaction effects (as can be seen on Figure 6), that is, /b/ vs. sonorants do not cause differences in voicing to the six classes that are examined here. Figure 7 illustrates how small the difference between the two groups is.

This suggests that RVA, which seems to be categorical in Slovak, is extended to the pre-sonorant position as well. The statistically significant difference observed is not due to less amount of voicing in obstruents before sonorant consonants as opposed to /b/, but rather to the fact that there are more instances when RVA fails to apply in pre-sonorant position (Figure 8). In these cases word-final devoicing occurs, which – when averaged across the board – gives a result between voiceless and voiced obstruents. This is in accordance with Strycharczuk (2012), who claims that pre-sonorant voicing is categorical but optional.

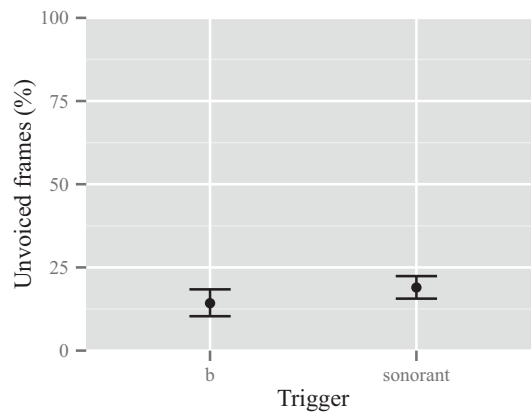


Figure 7.

Mean values for the voiceless–voiced ratio before /b/ and before sonorant consonants (error bars indicate 95% confidence intervals)

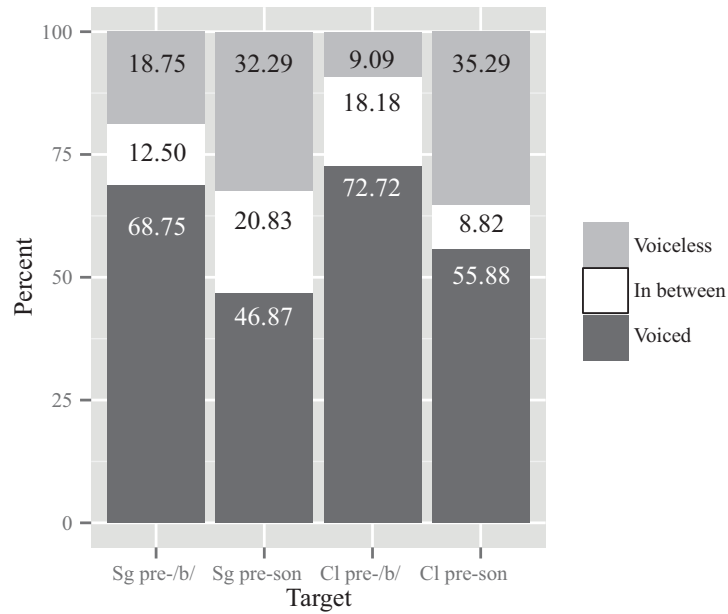


Figure 8.

Percentages of voiceless, partially voiced and voiced realizations of singleton obstruents and obstruent clusters in pre-/b/ and pre-sonorant position

We divided our data into three categories: (i) fully voiced realization, (ii) partially voiced realization and (iii) voiceless realization. The criteria for the classification were the following arbitrarily determined values. We classified those instances as “fully voiced” which contained up to 29% of unvoiced frames, that is they were at least 70% voiced. “Voiceless” occurrences were those which contained at least 71% of unvoiced frames, the rest were labelled as “in between”, that is, partially voiced. Figure 8 clearly demonstrates that it is not so much the partially voiced realizations that are considerably more numerous in the case of pre-sonorant obstruents but the instances of voiceless realizations. Therefore, in the remainder of this section, we will treat the pre-sonorant and pre-/b/ contexts as one group.

In Figure 9 we compare the amount of voicing between single consonants – this group comprises singleton obstruents and those single obstruents that remain from clusters with C₂ deletion – and CC clusters in pre-/b/ and pre-sonorant position. The two groups do not differ with regard to the voiced–voiceless portion during the obstruent(s): $b = -1.919$, $t(24) = -1.522$, $p = 0.141$.

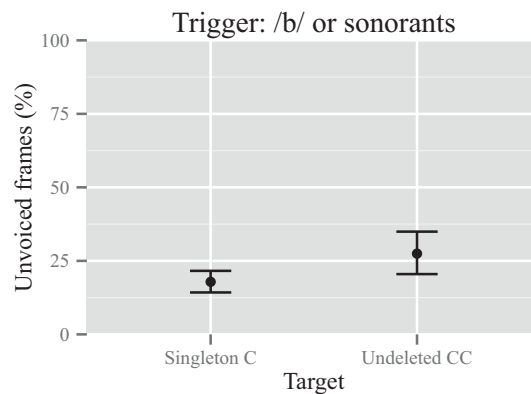


Figure 9.

Mean values for the ratio of the unvoiced part to total consonant length of single obstruents vs. obstruent clusters (error bars indicate 95% confidence intervals)

This result is indicative of a non-coarticulatory process since it is not the absolute voicing duration that is similar (see Figure 10), but the *proportion* of voicing. If the absolute voicing duration is the same or similar across different consonant lengths, we might suspect that voicing is due to articulatory inertia, so it is coarticulatory. If it is the proportion of voicing that is similar in single consonants and consonant clusters, it should probably be attributed to a pre-planned articulatory gesture. As Figure 10 shows, the duration of the

voiced portion of the obstruent cluster is significantly longer than the voiced portion of the single consonant, although the magnitude of the effect is very small [$b = -0.004$, $t(22) = -4.638$, $p = 0.0001$, $r = 0.001$].

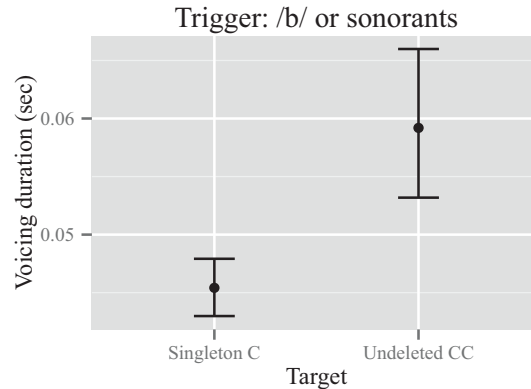


Figure 10.

Mean values for the duration of the voiced part in single obstruents vs. obstruent clusters in voicing context (error bars indicate 95% confidence intervals)

If we tease apart the single consonant class and compare true singletons with single consonants that result from cluster simplification in pre-consonant environment, we find that the two groups are not different with regard to their duration [$b = 2.99$, $t(10) = 1.352$, $p = 0.2062$], but they significantly differ in the voiced portion (again, however, the effect size is very small): $b = -0.003$, $t(24) = -3.879$, $p = 0.0007$, $r = 0.001$ (see Figure 11).

Interestingly enough, if we compare the voiced interval of reduced clusters with those of fully realized clusters we do not find a statistically significant difference [$b = -0.001$, $t(24) = -1.497$, $p = 0.147$], while they do differ with regard to their voicing ratio [$b = -1.958$, $t(24) = -2.98$, $p = 0.0065$, $r = 0.37$]. These results indicate that in the case of reduced stops there is a planned but unrealized cluster the voicing of which is implemented by speakers despite the deletion of C_2 . This finding suggests that RVA in Slovak is not coarticulatory, although, we must warn the reader that this experiment should be replicated with a larger set of data, which also includes non-alveolar stops so that cluster simplification may be avoided. In Figure 12 we sum up the effects of voicing on obstruents in Slovak.

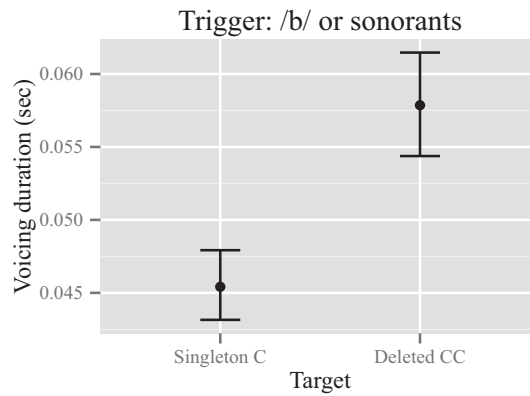


Figure 11.

Mean values for the duration of the voiced part in intended single obstruents vs. reduced obstruent clusters in voicing context (error bars indicate 95% confidence intervals)

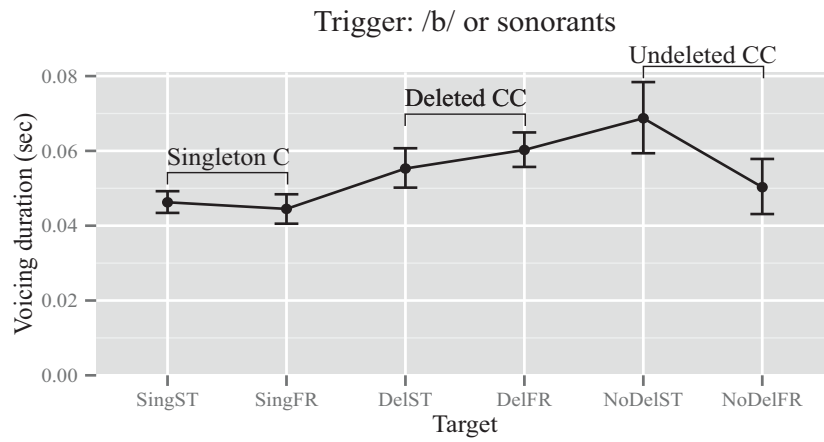


Figure 12.

Mean values for the duration of the voiced part in intended single stops and fricatives, reduced stop and fricative clusters and fully realized stop + stop and fricative + stop clusters in voicing context (error bars indicate 95% confidence intervals)

We can see in Figure 12 that single consonants are different from simplified and undeleted clusters, while the latter two are not statistically significant with regard to their voicing duration only their voicing ratio. Stops tend to be more voiced than fricatives. This tendency is more robust in the case of

longer intervals, that is, in undeleted clusters. It seems that fricatives reach a “voicing ceiling” earlier than stops, which is not unexpected due to the aerodynamic difficulty of initiating and maintaining voicing in fricatives as opposed to stops (e.g., Ohala 1983; Stevens 1998). As this paper has also shown, vowels are consistently longer before fricatives than before stops, which appears to be the case before single fricatives as well as fricative-initial clusters. We leave the investigation of this issue for future research.

A last piece of evidence we cite here supporting our claim that voicing assimilation in Slovak is not coarticulatory but rather categorical (and at times, optional), comes from Beňuš–Trnka (2014), who demonstrate that conversational fillers starting with a voiced schwa-like vowel like *umm*, *ur*, etc. function as prosodic breaks and as such induce word-final devoicing. However, in a non-negligible number of cases they do trigger voicing assimilation, indicating that speakers display a bimodal behaviour with a choice between two categorical options: they either produce word-final devoicing or they implement pre-sonorant voicing.

Conclusions

This paper has presented a modest contribution to the study of voicing assimilation in Slovak, a language for which instrumental/experimental phonetic and phonological research is lacking. More specifically, we have investigated the voicing properties of three-consonant clusters (CC#C), and how voicing assimilation affects them, an area of Slovak phonetics and phonology which has not received enough attention either. This paper is also a contribution to the study of pre-sonorant voicing, a topic of growing interest both empirically and theoretically. Our experiment has shown that word-final obstruent clusters in Slovak (just like singletons) are realized completely voiceless. This finding indicates that Slovak obstruents are categorically targetless for voicing in this position, which is claimed to be a pre-requisite of pre-sonorant voicing. It has been also shown that sonorant consonants and voiced stops do not differ in their voicing “capabilities” in this language, thus sonorants do not form an in-between category between voiceless obstruents and voiced obstruents. A novel finding of the paper is that pre-sonorant voicing assimilation in Slovak appears to be optional but categorical, rather than obligatory or gradual. We have not found any evidence for voicing assimilation, including pre-sonorant voicing, being a coarticulatory process for any of the speakers of our experiment. An unexpected result of the paper that is in need of further clarification is that vowels before fricatives are realized consistently longer than vowels before stops, irrespective of the prosodic position. It would also be interesting to study on a larger set of data whether C₂ deletion varies systematically with voicing.

References

- Bárkányi, Zsuzsanna – Kiss, Zoltán 2007. *A phonetically-based approach to the phonology of [v]: A case study from Hungarian and Slovak*. Paper presented at the 4th Old World Conference in Phonology. Rhodes, Greece. 18–21 January.
- Bárkányi, Zsuzsanna – G. Kiss, Zoltán 2012. *On the border of phonetics and phonology: Sonorant voicing in Hungarian and Slovak*. Paper presented at the 20th Manchester Phonology Meeting, Manchester. May 24–26th.
- Bárkányi, Zsuzsanna – G. Kiss, Zoltán 2013. *Phonetics or phonology? Why do sonorants not voice in Hungarian?* Paper presented at the 11th International Conference on the Structure of Hungarian, Piliscsaba. August 29–31.
- Beňuš, Štefan – Gafos, Admantios 2007. Articulatory characteristics of Hungarian ‘transparent’ vowels. *Journal of Phonetics* 35. 271–300.
- Beňuš, Štefan – Trnka, Marián 2014. *Prosody, voice assimilation and conversational fillers*. Manuscript. Institute of Informatics, Slovak Academy of Sciences.
- Boersma, Paul – Weenink, David 2012. Praat: Doing phonetics by computer. (Version 5.3.12) [Computer program]. Retrieved from <http://www.praat.org/>
- Bonet, Eulalia – Lloret, María-Rosa 1998. *Fonologia catalana*. Ariel, Barcelona.
- Burton, Martha – Robblee, Karen 1997. A phonetic analysis of voicing assimilation in Russian. *Journal of Phonetics* 25. 97–114.
- Charles-Luce, Jan 1985. Word-final devoicing in German: Effects of phonetic and sentential contexts. *Journal of Phonetics* 13. 309–324.
- Charles-Luce, Jan 1993. The effects of semantic context on voicing neutralization. *Phonetica* 50. 28–43.
- De Schutter, Georges – Taeldeman, Johan 1986. Assimilatie van Stem in de Zuidelijke Nederlandse Dialecten. In Devos, Maud – Johan Taeldeman (eds.): *Vruchten van z’n akker: opstellen van (oud-) medewerkers en oud-studenten voor Prof. V.F. Vanacker*. Seminarie voor Nederlandse Taalkunde, Ghent. 91–133.
- Dinnsen, Daniel A. 1985. A re-examination of phonological neutralization. *Journal of Linguistics* 21. 265–279.
- Dinnsen, Daniel A. – Jan Charles-Luce 1984. Phonological neutralization, phonetic implementation, and individual differences. *Journal of Phonetics* 12. 49–60.
- Field, Andy – Miles, Jeremy – Field, Zoe 2012. *Discovering statistics using R*. Sage, London.
- Fourakis, Marios – Iverson, Gregory 1984. On the ‘incomplete neutralization’ of German final obstruents. *Phonetica* 41. 140–149.
- Hallé, Pierre A. – Adda-Decker, Martine 2011. Voice assimilation in French obstruents: Categorical or gradient? In Goldsmith, John – Hume, Elizabeth – Wetzels, Leo (eds.): *Tone and features: phonetic and phonological perspectives*. De Gruyter, Berlin. 149–175.
- Hayes, Bruce 1999. Phonetically driven phonology: The role of Optimality Theory and inductive grounding. In Darnell, Michael – Moravcsik, Edith A. – Noonan, Michael – Newmeyer, Frederick J. – Wheatly, Kathleen (eds.): *Functionalism and formalism in linguistics*. Vol. 1. John Benjamins, Amsterdam–Philadelphia. 243–285.
- Hayes, Bruce – Steriade, Donca 2004. The phonetic bases of phonological markedness. In Hayes, Bruce – Kirchner, Robert M. – Steriade, Donca (eds.): *Phonetically based phonology*. Cambridge University Press, Cambridge. 1–33.

- Hualde, José-Ignacio 2005. *The sounds of Spanish*. Cambridge University Press, Cambridge.
- Hyman, Larry M. 1975. *Phonology: Theory and analysis*. Holt, Rinehart and Winston, New York.
- Jansen, Wouter 2004. *Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch*. Doctoral dissertation, Rijksuniversiteit Groningen.
- Jassem, Wiktor – Richter, Lutosława 1989. Neutralization of voicing in Polish obstruents. *Journal of Phonetics* 17. 317–325.
- Javkin, Hector R. 1976. The perceptual basis of vowel duration differences associated with the voiced/voiceless distinction. *Report of the Phonology Laboratory, UC Berkeley* 1. 78–92.
- Jiménez, Jesús – Lloret, María-Rosa 2008. Asimetrías perceptivas y similitud articulatoria en la asimilación de sonoridad del catalán. *Cuadernos de Lingüística del I.U.I. Ortega y Gasset* 15. 71–90.
- Kahlen-Halstenbach, Birthe 1990. Zur psychologischen Realität der Auslautverhärtung im Deutschen. *Zeitschrift für phonetische Sprachwissenschaft und Kommunikationsforschung* 43. 645–655.
- Kaplan, Abby 2011. How much homophony is normal? *Journal of Linguistics* 48/2. 1–41.
- Kingston, John – Diehl, Randy L. 1994. Phonetic knowledge. *Language* 70. 419–454.
- Kluender, Keith R. – Diehl, Randy L. – Wright, Beverly A. 1988. Vowel length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics* 16. 153–169.
- Markó, Alexandra – Grácz, Tekla Etelka – Bóna, Judit 2010. The realisation of voicing assimilation rules in Hungarian spontaneous and read speech: Case studies. *Acta Linguistica Hungarica* 57. 210–238.
- Massaro, Dominic W. – Cohen, Michael M. 1983. Consonant/vowel ratio: An improbable cue in speech perception. *Perception and Psychophysics* 33. 502–505.
- O'Dell, Michael – Port, Robert F. 1983. Discrimination of word final voicing in German. *Journal of the Acoustical Society of America* 73. Supplement 1.
- Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In MacNeilage, Peter (ed.): *The production of speech*. Springer, New York. 189–216.
- Parker, Ellen M. – Diehl, Randy L. – Kluender, Keith R. 1986. Trading relations in speech and non-speech. *Perception and Psychophysics* 39. 129–142.
- Pauliny, Eugén 1979. *Slovenská fonológia*. [Slovak phonology]. Slovenské Pedagogické Nakladateľstvo, Bratislava.
- Pinheiro, Jose – Bates, Douglas – DebRoy, Saikat – Sarkar, Deepayan 2013. The nlme Package. Linear and Nonlinear Mixed Effects Models. *R package version 3*. 1–108.
- Port, Robert F. – Dalby, Jonathan. 1982. C/V ratio as a cue for voicing in English. *Perception and Psychophysics* 2. 141–152.
- Port, Robert F. – Leary, Adam P. 2005. Against formal phonology. *Language* 81. 927–964.
- Port, Robert F. – Mitleb, Fares – O'Dell, Michael 1981. Neutralization of obstruent voicing in German is incomplete. *The Journal of the Acoustical Society of America* 70. Supplement 1.

- R Core Team 2012. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <http://R-project.org>.
- Recasens, Daniel – Mira, Meritxell 2012. Voicing assimilation in Catalan two-consonant clusters. *Journal of Phonetics* 40. 639–654.
- Recasens, Daniel – Mira, Meritxell 2013. Voicing assimilation in Catalan three-consonant clusters. *Journal of Phonetics* 41. 264–280.
- Rubach, Jerzy 1994. *The lexical phonology of Slovak*. Clarendon Press, Oxford.
- Rubach, Jerzy. 1996. Nonsyllabic analysis of voice assimilation in Polish. *Linguistic Inquiry* 27. 69–110.
- Silverman, Daniel. 2012. *Neutralization*. Cambridge University Press, Cambridge.
- Simon, Ellen 2010. Phonological transfer of voicing and devoicing rules. Evidence from L1 Dutch and L2 English conversational speech. *Language Sciences* 32. 63–86.
- Slowiaczek, Louisa M. – Dinnsen, Daniel A. 1985. On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics* 13. 325–341.
- Slowiaczek, Louisa M. – Szymanska, Helena J. 1989. Perception of word-final devoicing in Polish. *Journal of Phonetics* 17. 205–212.
- Steriade, Donca. 1997. *Phonetics in phonology: The case of laryngeal neutralization*. Manuscript. University of California Los Angeles.
- Steriade, Donca. 2008. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In Inkelas, Sharon – Hanson, Kristin (eds.): *The nature of the word*. MIT Press, Cambridge, MA. 151–180.
- Stevens, Kenneth N. 1998. *Acoustic phonetics*. MIT Press, Cambridge, MA.
- Strycharczuk, Patrycja. 2012. *Phonetics–phonology interactions in pre-sonorant voicing*. Doctoral dissertation, University of Manchester.
- Strycharczuk, Patrycja – Simon, Ellen 2013. Obstruent voice before sonorants. The case of West-Flemish. *Natural Language and Linguistic Theory* 31/2. 563–588.
- Wells, John C. 1982. *Accents of English 1–3*. Cambridge University Press, Cambridge.
- Wheeler, Max W. 2005. *The phonology of Catalan*. Oxford University Press, Oxford.
- Wright, Richard 2001. Perceptual cues in contrast maintenance. In Hume, Elizabeth – Johnson, Keith (eds.): *The role of speech perception in phonology*. Academic Press, San Diego. 251–277.
- Wright, Richard 2004. A review of perceptual cues and cue robustness. In Hayes, Bruce – Kirchner, Robert M. – Steriade, Donca (eds.): *Phonetically based phonology*. Cambridge University Press, Cambridge. 34–57.

The present research was supported by the Hungarian National Research Fund (OTKA K104897).

AZ ÚGYHOGY FUNKCIÓBŐVÜLÉSE A SPONTÁN BESZÉDBEN

Gyarmathy Dorottya

Bevezetés

A spontán beszédben a nyelvi jelek egy része többféle funkciót is betölthet; előfordulhatnak szintaktikai és nem szintaktikai pozícióban. Egyes elemek a közlésben az eredeti szintaktikai funkciójuk mellett töltelékszavakként, illetve diskurzusjelölőként is megjelenhetnek. A fonetikai és pszicholingvisztikai paradigmában egy elem akkor tekinthető töltelékszónak, ha megjelenésére az adott közlésben semmiféle grammatikai vagy szemantikai magyarázat nincsen, tartalmilag nem illeszkedik a közlésbe (Glücksmannová 2008), a tervezési diszharmóniák feloldására szolgál (vö. Gósy 2005; Gósy–Horváth 2009). A terminus tartalmilag nem áll szemben a *diskurzusjelölő*vel mint szakszóval, használata az adott diszciplína hagyományait követi.

Töltelékszói szerepben általában nyelvtől függetlenül funkciószavak fordulnak elő, mint például a német *also*, az angol *well* vagy a magyar *hát, tehát, így, izé*; jóllehet elvileg bármilyen szófajú szó válhat töltelékszóvá. Egyes nyelvekben előfordulnak olyan szavak vagy szókapcsolatok töltelékszói szerepben, amelyek tartalmas szavakat is magukban foglalnak, például az angol *you know* vagy a magyar *úgymond, aszongya*. Egy lexéma a töltelékszóvá válás során a spontán közlésekben elveszti a megszokott tartalmát, megváltozik az eredeti funkciója, de az új funkció mellett egyben a régi is használatban marad. A folyamat egyfajta szinkrón nyelvi változásnak tekinthető, amely akkor megy végbe, ha minél több beszélő kezdi használni az adott lexémát az új funkciójában (Pierrehumbert 2001; Bybee 2001). Fonetikai, illetve pszicholingvisztikai szempontból tehát egy elem töltelékszónak tekinthető akkor, ha a beszédben az eredeti szintaktikai funkciójától eltérő szerepet tölt be, például a hallgató számára a feldolgozáshoz, a beszélőnek a tervezéshez szükséges időt biztosítja (vö. Huszár 1985), a társalgásokban a beszélőnek a közlés folytatására irányuló szándékát jelzik (Schachter et al. 1991; Bortfeld et al. 2001), vagy a beszélgető felek segítségükkel koordinálják a társalgást és a beszélőváltásokat (Clark 1994). Pragmatikai szempontból azonban ezek a funkciók (beszédszándék jelzése, a beszélőváltások koordinálása) már egybeesnek a diskurzusjelölői szerepekkel.

Dér (2010) álláspontja szerint, ha egy nyelvi elem jól azonosítható nyelvi funkcióval bír, szintaktikai, szemantikai, grammatikai vagy pragmatikai-szövegszervezési szerepet tölt be, nem nevezhető töltelékelemnek. Ezekben

az esetekben diskurzusjelölőkről beszélhetünk. Ezek olyan – diskurzusszegmenseket összekapcsoló – kifejezések, amelyek elsősorban a kötőszók, határozószók és az előjárós szerkezetek szintaktikai osztályából származnak. Fő funkciójuk az általuk bevezetett S2 szegmens és az azt megelőző S1 szegmens interpretációja közti viszony jelzése (Fraser 1999). A diskurzusjelölők osztályát legszembetűnőbben a heterogenitás jellemzi; tagjai a legkülönbözőbb szófaji osztályokból származnak. Az egyszerűbbek igékre, határozószókra, kötőszókra, indulatszavakra, előjárószavakra vezethetők vissza; míg a komplexebbek akár egész szerkezeteket, tagmondatokat is magukba foglalhatnak. Közös jellemzőjük, hogy nincsenek egységes formai tulajdonságaik, ezért csak funkcionálisan adhatók meg. Szemantikailag jellemző rájuk, hogy a megnyilatkozás igazságfeltételeire általában nincsenek hatással, annak propozicionális tartalmát nem befolyásolják, de emocionális és expresszív funkcióval bírnak (Jucker 1993). A diskurzusjelölők nem csupán diskurzusszegmenseket kötnek össze – ezzel nagymértékben hozzájárulva a szöveg koherenciájának fenntartásához (Lenk 1998) –, de a közlésben határjelölő szerepet is betöltenek (vö. Jucker 1993; Fraser 1999; Fox Tree–Schrock 2002; Dér 2005, 2008; Markó–Dér 2008).

A magyar spontán beszédben folyamatosan szélesedik azon szavak köre, amelyek funkcióbővülésen mennek keresztül. Az elmúlt évek kutatásai számos lexéma – mint például az *ilyen* és a *tehát* (Horváth 2009), az *akkor* (Gósy 2009) és a *szóval* (Gósy 2013) esetében – igazolták a szinkrón nyelvi változásnak tekinthető folyamatot. Az említettekhez hasonló, egyes szavakra irányuló kutatások különös jelentőséggel bírnak, hiszen objektív akusztikai fonetikai vizsgálatokkal bizonyítják a szinkrón változást.

A jelen kutatás az *úgyhogy* funkcióbővülését kívánja igazolni. A spontán beszédben ugyanis egyre gyakrabban jelenik meg eredeti, szótározott jelentésétől eltérő funkcióban. Az *úgyhogy* alárendelő kötőszó oksági (következtető, illetve magyarázó) viszonyt fejez ki. A Magyar Értelmező Kéziszótár szerint: „ksz <Következményes m-ban:> ennek következtében, így” (Pusztai szerk. 2003: 1391). Az etimológiai kutatások tanúsága szerint grammatikalizációval keletkezett, melynek folyamán az *úgy* utalószó klitizálódott a *hogy* kötőszóhoz. Mindezt megelőzte a szórendi és a hangsúlyviszonyok megváltozása és a tagmondathatár eltolódása; a folyamatokat azonban nem kísérte alaki redukció (vö. Haader 2001; bővebben: Juhász 1992; Rácz 1995).

Az *úgyhogy* eredeti kötőszói használatától eltérő funkciójával már több tanulmány is foglalkozott. Markó és Dér (2008) egy négyfős társalgásban (114 perces anyagban) pragmatikai és akusztikai fonetikai szempontok alapján elemezte hat további diskurzusjelölővel (*így, meg, most, tehát, tényleg, vagy*) együtt azok szintaktikai és nem szintaktikai funkciójú használatát; míg Dér (2010) a *hát*, az *úgyhogy*, az *így* és az *ilyen* lexémák diskurzusjelölői megjelenését elemezte pragmatikailag. A jelen kutatás újdonsága abban rejlik, hogy nagy mennyiségű spontán narratívában elemzi az *úgyhogy* beszédben betöl-

tött funkcióit. Hipotézisünk szerint adataink alátámasztják az *úgyhogy* napjainkban zajló funkcióbővülését, és ez az artikulációs gesztusokban is megmutatkozik majd. Azt feltételezzük tehát, hogy a különböző nyelvhasználati szerepekben eltérően alakulnak az akusztikai paraméterek, hasonlóan ahhoz, ahogy azt a korábbi kutatások az *ilyen*, a *tehát* (Gósy–Horváth 2009) és az *izé* (Gyarmathy 2012) esetében is igazolták.

Anyag, módszer, kísérleti személyek

Az *úgyhogy* lexéma lehetséges funkciófüggő ejtésvariációit a BEA spon-tánbeszéd-adatbázis (Gósy et al. 2012) 20 felvételében vizsgáltuk. A 10 női és 10 férfi beszélővel rögzített hanganyagokból a spontán beszédet tartalmazó felvételrészeket (narratíva, véleménykifejtés, interpretált beszéd) elemeztük. Adatközlőink egynyelvű, budapesti, köznyelvet beszélő felnőttek; átlagéletkoruk 41,95 év. A 20 beszélő mindegyikénél tudtunk *úgyhogy*-előfordulást ad-atolni; az összesen 5 óra 48 percnyi hanganyagban 155 darabot, ami szemé-lyenként átlagosan 7,75 darab lexémát jelent. A spontán beszédben tehát hoz-závetőlegesen mintegy 2,25 percenként fordul elő az *úgyhogy*. Az egyes előfordulásokat először kigyűjtöttük a hanganyagból, majd a Praat szoftver 5.1.34-es verziójával (Boersma–Weenink 2009) annotáltuk: meghatároztuk az időtartamukat, az f_0 -átlagukat, megvizsgáltuk prozódiai függetlenségüket. Az utóbbi esetben azt elemeztük, hogy az *úgyhogy*-ot megelőzően és/vagy köve-tően megfigyelhető-e szünettartás, ezek milyen időtartamúak, néma vagy ki-töltött szünetként (esetleg ezek kombinációiként) realizálódnak-e. A szó idő-tartamát a kezdő magánhangzó első hangszalagrezgésétől a záró mássalhang-zó lecsengéséig mértük. Az alaphangmagasság értékeit automatikusan nyer-tük ki oly módon, hogy a Praat programmal minden 10 ms-ban kiírtattuk az f_0 -értéket. Az automatikus alaphangmagasság-kinyerésnél 150 ms-os Ham-ming-típusú ablakkal dolgoztunk a következő tartományokban: a férfiak ese-tében 60–350 Hz, a nőknél 75–450 Hz. Az f_0 -görbe jobb reprezentálásának érdekében interpolációt és 5 pontos mediánszűrést alkalmaztunk. Mindezt auditív és vizuális ellenőrzés követte; a kapott értékek közül kiszűrtük a mé-rési hibából adódó kiugró szélső értékeket. Meghatároztuk az *úgyhogy* teljes időtartamára számolható átlagos alaphangmagasságot, az összehasonlító elemzésekben ezeket az átlagértékeket használtuk (a hanglejtésváltozatokat nem vettük figyelembe). A példák kategorizálásához meghatároztuk a köz-lésbeli szerepüket és tagmondatbeli helyzetüket. Elsőként az eredeti kötőszói használatot különítettük el az ettől eltérő funkciójú előfordulásoktól, majd az utóbbiaknál kontextuselemzés alapján meghatároztuk a közlésben betöltött szerepüket. Az adatok statisztikai elemzését párosított t -próbával végeztük 95%-os konfidenciaintervallumon, amelyet minden esetben normalitásvizsgá-lat előzött meg (Shapiro–Wilk-teszt). A statisztikai elemzéshez az SPSS 19.0 verziószámú szoftvert használtuk.

Eredmények

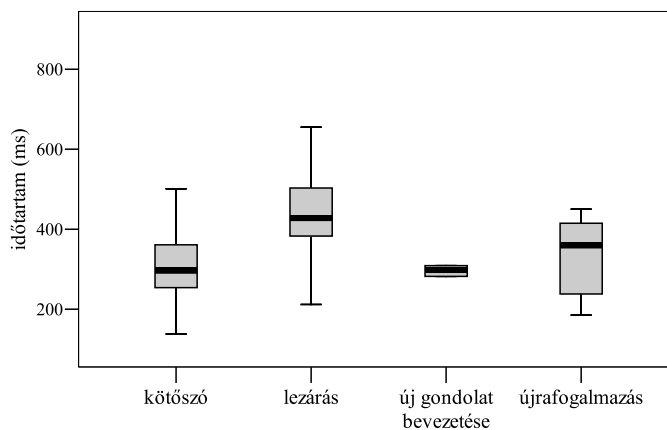
A 155 darab *úgyhogy*-előfordulást elsőként a közlésben betöltött funkciójuk szerint osztályoztuk. A kategorizálás során kontextuselemzést végeztünk, amely döntően a Dér (2010) által bevezetett kritériumrendszeren alapul; kiegészítve az általunk felvett új kategóriával (új gondolat bevezetése). Dér (2010) három funkciót különít el: (i) következtető vagy magyarázó viszonyt kifejező alárendelő **kötőszó** (ez minden esetben egy tagmondatot előz meg); (ii) **újrafogalmazás**, amikor a beszélő a már elhangzott közlését a megértés pontosításának céljából átfogalmazza; és (iii) **lezárás**, amikor virtuális mondat végén, az adott gondolatmenet zárásaként, nem ritkán a szóátadási szándék jelzésére szolgál. Az általunk elemzett példák szükségessé tették egy negyedik kategória bevezetését is: (iv) a beszélők néhány esetben egy teljesen **új**, a korábbiakhoz szervesen nem kapcsolódó **gondolatot** vezettek be az *úgyhogy* szóval. Ezt a kategóriát aszerint lehet elhatárolni az előzőtől, hogy amíg abban az esetben az *úgyhogy* a megnyilatkozás végén fordul elő, záró elemként, addig az ebbe a kategóriába tartozó *úgyhogy*-ok kivétel nélkül a következő megnyilatkozás bevezető elemeként a közlésegység elején jelennek meg. Ezt a legtöbb esetben egy hosszabb megelőző szünettartás is jellemzi (l. alább).

A példák 53,55%-ában az *úgyhogy* szintaktikai funkcióban, tehát kötőszóként fordult elő, például: *már pedig a házasságok többnyire tönkre szoktak menni több mint tíz éve váltam úgyhogy tudom én ezt*. Az esetek 46,45%-ában azonban az adatközlők nem szintaktikai funkcióban alkalmazták a szócskát. 39,35%-ban lezáró szerepben fordult elő (*meg volt két protokoll-jegyünk akkor már miért ne menjünk úgyhogy*); 3,87%-ban új gondolatot vezetett be [*egyre följebb csúszik a hóhatár az Alpokban meg a Kárpátokban is* (néma szünet 1479 ms) *úgyhogy*öő (124 ms) *a másik kérdésről meg azt gondolom hogy...*]; míg 3,23%-ban újrafogalmazást jelölt (*rengeteg virágom van a lakásban úgyhogy minden szoba tele van virággal*). Az arányszámokból jól látható, hogy a hagyományos, kötőszói használat mellett a beszélők a spontán beszéd során már csaknem ugyanolyan mértékben alkalmazzák az újabb funkciókban is az *úgyhogy*-ot. Ebben a használati arányban a diskurzusjelölővé válás folyamata érhető tetten.

Elemeztük az *úgyhogy* **tagmondatbeli helyzetét**. A példák 60%-a tagmondat elején, 40%-a tagmondat végén fordult elő. A kép tovább árnyalható, ha a szintaktikai funkcióban való előfordulás mentén folytatjuk elemzésünket. A szintaktikai funkcióban (kötőszóként) megjelenő *úgyhogy*-ok szükségszerűen kivétel nélkül tagmondat elején fordultak elő; a nem szintaktikai funkcióban használtak túlnyomó többsége (84,72%) ezzel szemben tagmondat végén, kisebb hányaduk (15,28%) pedig tagmondat elején volt adatolható. Alapvetően tehát a tagmondatbeli helyzet alapján is el lehet különíteni egymástól az *úgyhogy* szintaktikai és nem szintaktikai funkciójú használatát. A nem szintaktikai szerepű *úgyhogy* három kategóriájában a következőképpen alakult a tag-

mondatbeli helyzet. A lezáró *úgyhogy*-ok természetesen kivétel nélkül tagmondat-, illetve megnyilatkozásvégen voltak adatolhatók, míg az új gondolatot bevezetők, illetőleg az újrafogalmazásként használtak mindegyike – magától értetődően – tagmondatok, megnyilatkozások elején.

Elemeztük az egyes *úgyhogy*-előfordulások **időtartamát** a négy funkcióban. Kötőszóként átlagosan 316,4 ms (átlagos eltérés: 99,9 ms), virtuális mondat lezárásaként 447,2 ms (átlagos eltérés: 109,8 ms), új gondolat bevezetésekor 286,8 ms (átlagos eltérés: 83,8 ms), míg újrafogalmazás esetén 329,6 ms (átlagos eltérés: 114,0 ms) időtartamban realizálódnak (1. ábra).



1. ábra

Az *úgyhogy* időtartama az egyes funkciók szerint

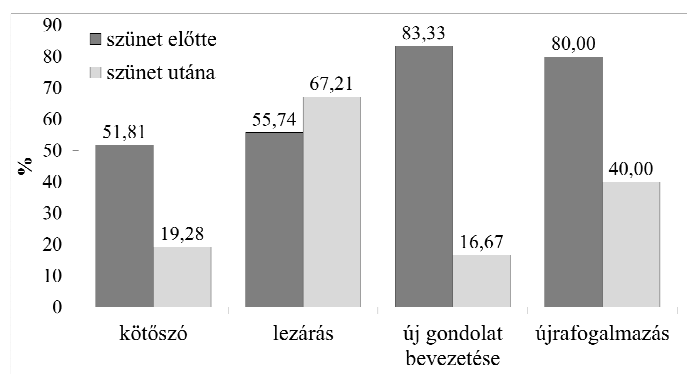
A beszélők tehát lezáró funkcióban ejtik a leghosszabban az *úgyhogy*-ot, ezzel a lassítással is mintegy jelezve beszédpartnerüknek, hogy a gondolamenetük végéhez értek. Amikor azonban új gondolatot vezetnek be ezzel a szóval, lerövidül az ejtés időtartama. A legrövidebb átlagos időtartam tehát tagmondat eleji, a leghosszabb pedig tagmondat végi pozícióban adatolható, ami összefüggésben állhat azzal a szakirodalomban már igazolt ténnyel, hogy a megnyilatkozások végén lassul az artikulációs tempó, a beszédhangok időtartama megnövekszik (vö. Magdics 1966; Kassai 1979; Berkovits 1993; Szaszák 2008; White–Mády 2008; Fletcher 2010).

Az adatok statisztikai elemzéséhez a kapott időtartamértékeket elsőként az egyes beszélőkhöz normalizáltuk, tehát minden beszélő esetében az általa ejtett *úgyhogy*-ok átlagidőtartamához. Ahogy azt korábban ismertettük, új gondolat bevezetésére, illetve újrafogalmazás céljából adatközlőink csak ritkán használták az *úgyhogy* szócskát, ezért ez a két csoport nagyon csekély elemszámmal bír (6, illetve 5 db); a további statisztikai elemzésekből ezeket kizár-

tuk. A két legnagyobb csoport (kötőszó és lezárás) értékeit az egyes beszélőkön belül vetettük össze. Megvizsgáltuk, hogy ugyanazon adatközlő milyen időtartamban ejti az *úgyhogy*-ot a két fő funkcióban. Tekintve, hogy anyagunkban a beszélők nem egyforma arányban ejtették a vizsgált lexémát az egyes funkciókban, akadt olyan adatközlő, akinél csak kötőszóként, és olyan is, akinél csak lezárásként tudtuk adatolni. Az összehasonlító vizsgálatba ezeknek a beszélőknek az adatait nem tudtuk bevonni, így 6 személytől összesen 20 példát voltunk kénytelenek kizárni az elemzésből. A kizárt példák közül 2 esetben (10%) lezárásként fordult elő az *úgyhogy*, míg 18 esetben kötőszóként (80%). Amely adatközlők beszédmintáiban tehát az *úgyhogy* csak egyféle szerepkörben jelenik meg, az eredeti, kötőszói használat az elsődleges. A két funkció időtartam-különbsége itt is megfigyelhető: kötőszóként átlagosan 292,4 ms-ban (átlagos eltérés: 95,8 ms), lezáró elemként pedig mintegy 100 ms-mal hosszabban, 391,5 ms-ban (átlagos eltérés: 60,1 ms) realizálódik. Azoknál a beszélőknél, akik mind szintaktikai, mind megnyilatkozás-záró funkcióban alkalmazzák a vizsgált lexémát, másként alakul a használati gyakoriság. A példák többsége (52,4%) ugyan itt is a kötőszói használatból adódik, az *úgyhogy* közlészáró lexémaként ebben a csoportban csaknem ugyanilyen gyakori (47,6%). A 14 adatközlő közül 4-nél ugyanolyan arányban (50-50%) találtunk példát a szintaktikai funkciójú és a diskurzusjelölői használatra; 4-üknél a kötőszói, míg 6-uknál a lezárói szerep volt a gyakoribb. Az elemzésbe bevont 14 beszélőnél az *úgyhogy* kötőszóként átlagosan 323,1 ms (átlagos eltérés: 100,7 ms), lezárásként pedig 449,1 ms (átlagos eltérés: 110,9 ms) hosszú volt. Az összehasonlító statisztikai elemzést megelőző normalitásvizsgálat szerint adataink normál eloszlásúak voltak, ezért párosított *t*-próbával elemeztük azokat, amely szerint az egyes beszélők megnyilatkozás-záró funkcióban szignifikánsan hosszabban ejtik az *úgyhogy*-ot, mint kötőszóként [$t(13) = -3,68$; $p = 0,003$].

Elemeztük az *úgyhogy* **prozódiai függetlenségét** a négyféle funkcióban, tehát azt, hogy a beszélők tartanak-e előtte, illetőleg utána (néma vagy kitöltött) szünetet, és milyen arányú a megelőző és követő szünetek előfordulása az egyes funkciókban. Előzetes elvárásunk az volt, hogy a nem szintaktikai funkciót betöltő *úgyhogy*-ok nagyobb prozódiai függetlenséggel bírnak; több tanulmány is a diskurzusjelölők közös vonásaként említi, hogy azok elkülönülnek a mondat többi részétől, amit a beszédben az adott lexéma előtt és után tartott szünet, mondatban a központosítás jelez (vö. Schourup 1999; Fraser 1999). Az általunk vizsgált anyagban ez a feltevés némiképp igazolódni látszik, hiszen ahogy az a 2. ábráról leolvasható, a nem szintaktikai szerepű *úgyhogy*-ok esetében nagyobb arányú szünettartás volt adatolható. Amikor az *úgyhogy* kötőszóként jelent meg, az esetek 51,8%-ában előzte meg szünet [*betettek a Petőfi szótárnak* (néma szünet 1019 ms) *úgyhogy az megcsökkent*], de csupán 19,3%-ában követte [*úgy telepítették ki annak idején Szlovákiából úgyhogy* (néma szünet 683 ms) *neki azért vannak éles emlékképei*]. Ahol

azonban az *úgyhogy* diskurzusjelölőként fordult elő, a példák 59,7%-át előzte meg, 61,1%-át követte néma vagy kitöltött szünet. A három diskurzusjelölői szerepkörben ez az arány az alábbiak szerint alakul: a gondolatmenet lezárásaként ejtett *úgyhogy*-oknál a beszélők az esetek 55,7%-ában tartanak előtte szünetet [*az asztaldíszben abban van élő fenyő* (néma szünet 1021 ms) **úgyhogy**], 67,2%-ában utána [*nem értettem miért az államvizsgán kell előjönnie ilyen attitűdöknek a tanszékvezető részéről úgyhogy* (340 ms) (néma szünet 638 ms) *ja és hát a témavezetőmmel pedig gyakorlatilag nem konzultáltam*]. Ha új gondolatot vezetnek be (83,3%) az *úgyhogy*-gyal [*ahol kereszteltek ott megyek férjhez* (néma szünet 1088 ms) *ööm* (514 ms) (néma szünet 836 ms) **úgyhogy** *tegnap elintéztük az éttermet*], illetve újrafogalmazzák (80,0%) korábbi mondandójukat [*szeretjük nem azért esszük mert újévkor lencsét kell enni* (néma szünet 333 ms) **úgyhogy** *szeretjük a lencsét*], csaknem minden esetben megfigyelhető előtte néma vagy kitöltött szünet. A követő szünetek az újrafogalmazás esetén a példák 40,0%-át, az új gondolat bevezetésekor 16,7%-át jellemzik. A kapott adatok némiképp ellentmondanak a korábbi vizsgálati eredményeknek, ahol nem találtak lényeges eltérést az *úgyhogy* diskurzusjelölői és szintaktikai funkciójú megjelenése között a prozódiai függetlenség tekintetében (vö. Markó–Dér 2008), amit az alkalmazott módszer és az adatközlők számának különbsége magyarázhat.

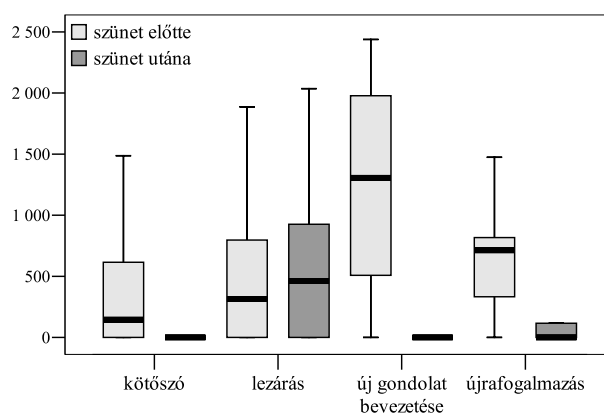


2. ábra

Az *úgyhogy*-ot megelőző és követő szünetek aránya

Elemeztük a megelőző és követő szünetek időtartamát az egyes beszédbeli funkciók szerint. Az eredmények alapján ebben a paraméterben is elkülönülnek egymástól a szintaktikai és a diskurzusjelölői szerepű *úgyhogy*-ok (3. ábra). A kötőszóként ejtett *úgyhogy*-ok előtt az adatközlők átlagosan 313,7 ms (átlagos eltérés: 381,7 ms), utána 142,3 ms (átlagos eltérés: 385,2 ms) hosszúságú szünetet tartanak. A diskurzusjelölői funkciókban – a megelőző szünetek

legalábbis – jóval hosszabb időtartammal realizálódnak. A leghosszabban akkor, ha új gondolatot vezetnek be (1256,2 ms; átlagos eltérés: 907,8 ms), körülbelül fele ilyen hosszán újrafogalmazás esetén (667,4 ms; átlagos eltérés: 555,1 ms) és megnyilatkozás lezárásakor 457,7 ms-mal (átlagos eltérés: 535,5 ms). A követő szünetek a kötőszói szerepű *úgyhogy*-okhoz hasonlóan lényegesen rövidebbek: új gondolat bevezetésekor átlagosan 134,8 ms (átlagos eltérés: 274,1 ms), újrafogalmazáskor 99,0 ms (átlagos eltérés: 164,4 ms) hosszúak. Az, hogy az említett két kategóriánál az adatközlők jellemzően inkább az *úgyhogy*-ot megelőzően tartanak szünetet, és azok mintegy 7–10-szer hosszabban realizálódnak, mint a követő szünetek, a beszédtervezés közbeni bizonytalanságra utal. Ilyenkor ugyanis a beszélők mindenképpen folytatni kívánják megnyilatkozásukat, maguknál akarják tartani a szót, a közlés tartalmi részében azonban még bizonytalanok, aminek átgondolásához időre van szükségük. A lezáró funkciójú *úgyhogy*-ok eredményei nem csupán a szintaktikai funkciójúaktól különböznek, de a diskurzusjelölői szerepek között is eltérő mintázatot mutatnak. Ennél a kategóriánál valamivel nagyobb volt a követő szünetek aránya, mint a megelőzőké, és az előzőek átlagosan hosszabbak is voltak: a megelőző szünetek átlagos időtartama: 457,7 ms (átlagos eltérés: 535,5 ms), a követő szünetek átlagos időtartama: 545,3 ms volt (átlagos eltérés: 524,4 ms).



3. ábra

Az *úgyhogy*-ot megelőző és követő szünetek időtartama

A temporális paraméterek alapján úgy tűnik, hogy ha egy skálán képzel-nénk el a diskurzusjelölővé válást az *úgyhogy* esetében, a skála kiindulópontja az eredeti, kötőszói használat, végpontja az ettől teljesen elkülöníthető megnyilatkozást lezáró szerep lehetne. A kapott mintázatok alapján a másik

két diskurzusjelölői használat még nem távolodott el annyira az eredeti szerepkörétől, hiszen a beszélők ezek után is (csakúgy, mint a kötőszó esetében) folytatni kívánják a közlést. Ezzel szemben lezáráskor gyakran nemcsak az aktuális virtuális mondatot, de az egész közlésrészt is befejezik, és átadják a társalgópartnerüknek a szót. A jelen kutatásban elemzett 61 megnyilatkozás-záró *úgyhogy* után a beszélők az esetek többségében (67,2%) csak az adott megnyilatkozást zárták le, de a közlésüket ezután folytatták [*hát most majd alkalmazni fogom ott úgyhogy* (néma szünet 517 ms) *jó szerintem ennyi ebből elég lesz akkor most beszéljünk egy kicsit más témáról*]. Mindössze a példák 32,8%-ában került sor az egész közlésrész lezárására, és szóátadásra.

Az *úgyhogy*-ot megelőző és követő szünetek időtartamát a statisztikai vizsgálatok elvégzéséhez először a csoportátlagokhoz és a szóráshoz normalizáltuk. Az elemzésbe természetesen ezúttal sem vontuk be a két legkisebb elemszámú csoportot (új gondolat bevezetése és újfogalmazás). A két legnagyobb elemszámú csoportban a megelőző és követő szünetek időtartamát az egyes beszélőkön belül vetettük össze, tehát azt vizsgáltuk, hogy az *úgyhogy* előtt, illetve után tartott szünetek hossza különbözik-e attól függően, hogy a lexémát kötőszóként vagy diskurzusjelölőként alkalmazzák. Ebben az elemzésben csak azoknak az adatközlőknek az adatait tudtuk figyelembe venni, akiknél mindkét funkcióban előfordult az *úgyhogy*. 6 személyt (20 adat) így kizártuk a vizsgálatból. A kizárt példáknaál a kötőszóként ejtett *úgyhogy*-okat átlagosan 285,1 ms (átlagos eltérés: 331,4 ms) hosszúságú szünet előzte meg, és 59,9 ms-os (átlagos eltérés: 254,1 ms) követte; míg a lezáró szerepűeket átlagosan 309,5 ms-os (átlagos eltérés: 437,7 ms) előzte meg, és 560,5 ms-os (átlagos eltérés: 77,1 ms) követte. Az elemzésbe bevont beszélők a szintaktikai funkciójú *úgyhogy*-okat megelőzően 329,1 ms (átlagos eltérés: 395,4 ms) hosszú szünetet tartottak, míg utána fele ilyen hosszút, 165,2 ms-osat (átlagos eltérés: 412,9 ms). A megnyilatkozást lezáró *úgyhogy*-oknál más mintázatot kapunk: a megelőző szünetek átlagos hossza 462,8 ms (átlagos eltérésük: 540,9 ms), a követőké 544,7 ms (átlagos eltérés: 533,2 ms). A két csoporton elvégzett normalitásvizsgálat mind a megelőző, mind a követő szünetek esetén normál eloszlást igazolt, így adatainkat párosított *t*-próbával elemeztük. Az eredmények szerint a megelőző szünetek hossza nem, de a követő szüneteké szignifikánsan különbözik egymástól aszerint, hogy a beszélő kötőszóként vagy a megnyilatkozás lezárásaként ejtette-e az *úgyhogy*-ot [$t(13) = -3,832$; $p = 0,002$].

Az **alaphangmagasságot** az *úgyhogy*-előfordulások teljes időtartamára vetítve elemeztük mindkét nemnél. A nőknél gyakorlatilag nem volt különbség a szintaktikai funkciójú (átlag: 187,6 Hz) és a diskurzusjelölői *úgyhogy*-ok (átlag: 190,4 Hz) f_0 -átlaga között, és a férfiaknál is csupán mintegy 10 Hz-nyi eltérést adatoltunk (kötőszó: 118,8 Hz; diskurzusjelölő: 128,9 Hz). Az adatok alapján tendenciaszerűen elmondható, hogy diskurzusjelölőként az *úgyhogy*-ot a beszélők magasabb alaphangon ejtik, mint az eredeti, kötőszói szerep-

ben. Ha a három diskurzusjelölői szerepben külön elemezzük az alaphangmagasságot, nem tapasztalunk lényeges eltérést az egyes csoportok között (az egyes beszédbeli funkciók részletes adatai az 1. táblázatban olvashatók). Itt érdemes megjegyezni, hogy az elemzett példák között csupán két esetben fordult elő irreguláris zöngé, egyszer egy kötőszóként, egyszer pedig egy lezárásként ejtett *úgyhogy*-nál. Ez a kisszámú előfordulás azért meglepő, mert az elmúlt években több nyelvre is – így a magyarra is (vö. Markó 2013) – számos kutatás igazolta a glottalizáció határjelző szerepét frázis-, illetve megnyilatkozásvégen (Lehiste 1965; Henton–Bladon 1988; Fant–Kruckenberg 1989; Slifka 2006)

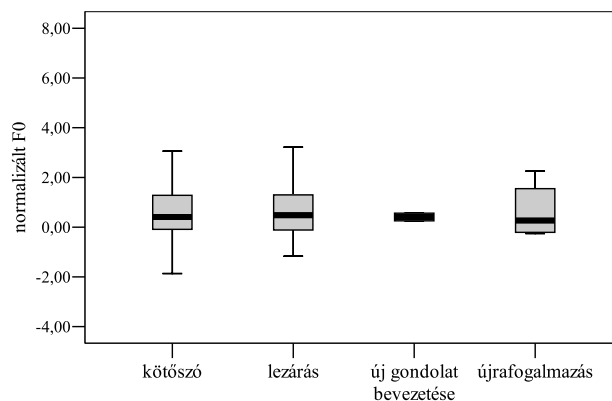
1. táblázat: Az *úgyhogy*-okban mért átlagos f_0 a négy funkcióban

Adatközlők	f_0 -átlag (Hz)			
	Kötőszó	Lezárás	Új gondolat bevezetése	Újrafogalmazás
Férfiak	118,8	128,9	128,6	–
Nők	187,6	189,9	189,0	195,5

Ahhoz, hogy a nők és a férfiak adatait együttesen tudjuk vizsgálni, az *úgyhogy*-okra kapott f_0 -értékeket az egyes beszélők átlagos alaphangmagasságához normalizáltuk; ezzel mind a nemekből adódó különbségeket, mind pedig az egyéni eltéréseket kiküszöböltük. A normalizált alaphangmagasság megmutatja, hogy az egyes *úgyhogy*-előfordulásokban mért f_0 -érték az adott beszélő átlagos alaphangmagasságánál magasabb vagy alacsonyabb-e. Ettől függően a kapott érték pozitív vagy negatív előjelű lehet. Az adatok szerint az *úgyhogy* mind a négy elemzett funkcióban a beszélők átlagos alaphangjához képest magasabb f_0 -lal realizálódik. Kötőszói szerepben 0,8 Hz (átlagos eltérés: 1,7 Hz), lezárásként 0,7 Hz (átlagos eltérés: 1,4 Hz), új gondolat bevezetésekor 0,5 Hz (átlagos eltérés: 0,8 Hz) és újrafogalmazáskor 0,7 Hz (átlagos eltérés: 1,1 Hz) az eltérés (4. ábra).

A statisztikai elemzéseknél a normalizált alaphangmagasság-értékeket szintén az egyes beszélőkön belül vetettük össze. A vizsgálatból ezúttal is kizártuk a két legkisebb elemszámú csoportot, illetve azon beszélők adatait, akiknél vagy csak kötőszói, vagy csak lezáró *úgyhogy*-ra találtunk példát. Az utóbbiaknál a kötőszóként ejtett *úgyhogy*-okban mért alaphangmagasság 2,4 Hz-cel (átlagos eltérés: 2,4 Hz) volt magasabb a beszélők átlagos alaphangjánál, míg a virtuális mondatot záróké csupán 0,2 Hz-cel (átlagos eltérés: 0,9 Hz). Az elemzésbe bevont beszélőknél az *úgyhogy* kötőszóként mindössze 0,4 Hz-cel (átlagos eltérés: 0,8 Hz), lezárásként pedig 1,1 Hz-cel (átlagos eltérés: 1,8 Hz) volt magasabb, mint az adatközlők átlagos alaphangmagassága. A két csoporton előzetesen normalitásvizsgálatot végeztünk, mely szerint adataink normál

eloszlásúak voltak. A párosított t -próba nem mutatott szignifikáns különbséget a két funkció között.



4. ábra

Az *úgyhogy*-okban mért f_0 -átlag a beszélők átlagos f_0 -jához normalizálva

Következtetések

A jelen kutatásban az *úgyhogy* lexéma akusztikai fonetikai vizsgálata alapján igyekeztünk választ találni arra a kérdésre, hogy az eredeti kötőszói használatán túl milyen egyéb funkciókat tölt be az *úgyhogy* a spontán beszédben, és az eltérő funkciók az akusztikumban is elkülöníthetők-e egymástól. A mintegy hatórányi spontán beszédből származó *úgyhogy*-előfordulásokat elsőként aszerint csoportosítottuk, hogy szintaktikai funkcióban vagy diskurzusjelölőként fordultak-e elő. Az adatok mintegy fele-fele arányban oszlottak meg a két kategória között. A diskurzusjelölői *úgyhogy*-okat további három csoportra lehetett osztani: megnyilatkozást lezáró, új gondolatot bevezető, illetve újrafogalmazó szerepet betöltőkre; az utóbbi kettőre csak néhány példát találtunk. Adataink alapján megállapítható, hogy a spontán beszédben az *úgyhogy* leggyakrabban kötőszóként (eredeti funkciójában), illetve megnyilatkozást záró elemként használatos, ami egyértelműen jelzi a kötőszó funkcióbővülését, diskurzusjelölővé válását.

A temporális elemzésekből kiderült, hogy az új gondolatot bevezető *úgyhogy*-ok a legrövidebbek, a lezáró szerepűek a leghosszabbak, míg kötőszói és újrafogalmazói funkcióban közel azonos időtartammal realizálódnak. A statisztikai elemzést csak a két legnagyobb elemszámú csoportra tudtuk elvégezni; az eredmények szerint matematikailag is igazolható a kötőszói és a lezáró funkciójú *úgyhogy*-ok időtartam-különbsége. A spontán beszédben betöltött szerep tehát meghatározza az ejtés időtartamát. Eredményeink egybevágóak Markó és Dér (2008) korábbi kutatásával; az *úgyhogy* diskurzusjelölő

lői előfordulásaiban mérhető hosszabb időtartam tagoló funkcióval bír, és egyben megfelelő időt biztosít a beszélő számára – amennyiben a partnere nem veszi át a szót – a folytatás megtervezésére.

A prozódiai függetlenség vizsgálatánál a megelőző és követő szünetek gyakoriságának mintázatai alapján a megnyilatkozást záró *úgyhogy*-ok csoportja egyértelműen elkülönült a többitől. Kizárólag erre a csoportra jellemző, hogy a beszélők többször tartanak szünetet az *úgyhogy* után, mint előtte. A másik három csoportnál a megelőző szünetek aránya volt a nagyobb, a szintaktikai és diskurzusjelölői funkció azonban itt is elkülönülni látszik: az utóbbinál jóval magasabb százaléku volt a megelőző szünetek előfordulása. Ehhez a mintázathoz illeszkedik a megelőző és a követő szünetek időtartama is a négy elemzett funkcióban. A diskurzusjelölői szerepekben a beszélők jóval hosszabb szünetet tartanak az *úgyhogy* előtt, mint amikor kötőszóként alkalmazzák azt, az új gondolatot bevezető és az újrafogalmazásként használt *úgyhogy*-ok azonban a követő szünetek időtartamában a forrásszófajra hasonlítanak. A lezáró szerepű *úgyhogy*-ok ebben a tekintetben is különböznek a többitől, esetükben a követő szünetek nem csak gyakoribbak, de hosszabban is realizálódnak, mint a megelőzők. A diskurzusjelölői *úgyhogy*-ok közül a legélesebben tehát a megnyilatkozást záró *úgyhogy* különül el az eredeti, kötőszói használatától. A két legnagyobb csoporton elvégzett statisztikai elemzés a követő szünetek hosszában igazolt szignifikáns különbséget.

Az *úgyhogy*-előfordulások teljes időtartamára vetített alaphangmagasság elemzése – előzetes elvárásainkkal ellentétben – nem mutatott ki különbséget a szintaktikai és a diskurzusjelölői funkciók között sem az objektív, sem a normalizált értékek tekintetében.

A kutatás eredményei igazolták egy, a jelenben zajló szinkrón nyelvi változás folyamatát az *úgyhogy* kötőszó esetében. A korpuszalapú vizsgálatból kiderült, hogy az *úgyhogy* egyre több funkciót tölt be a spontán beszédben, és az egyes szerepkörök a szó akusztikai vetületét is meghatározzák.

Irodalom

- Berkovits, Rochele 1993. Utterance-final lengthening and the duration of final-stop closures. *Journal of Phonetics* 21/4. 479–489.
- Boersma, Paul – Weenink, David 2009. *Praat: doing phonetics by computer* (Version 5.1.34) [Computer program]. <http://www.praat.org> (A letöltés ideje: 2011. október 2.)
- Bortfeld, Heather – Leon, Silvia D. – Bloom, Jonathan E. – Schober, Michael F. – Brennan, Susan E. 2001. Disfluency rates in conversations: Effects of age, relationship, topic, role and gender. *Language and Speech* 44/2. 123–147.
- Bybee, Joan L. 2001. *Phonology and language use*. Cambridge University Press, New York.
- Clark, Herbert H. 1994. Managing problems in speaking. *Speech Communication* 15. 243–250.

- Dér Csilla Ilona 2005. Diskurzusszerveződés és grammatikalizáció. *Nyelvtudományi Közlemények* 102. 247–264.
- Dér Csilla Ilona 2008. Mik is a diskurzusjelölők? In Keszler Borbála – Balázs Géza (szerk.): *Diskurzus a grammatikában – grammatika a diskurzusban*. Tinta Kiadó, Budapest. 293–303.
- Dér Csilla Ilona 2010. „Töltelékelem” vagy új nyelvi változó? A hát, úgyhogy, így és ilyen újabb funkciójáról a spontán beszédben. *Beszédkutatás* 2010. 159–170.
- Fant, Gunnar – Kruckenberg, Anita 1989. Preliminaries to the study of Swedish prose reading and reading style. *Speech Transmission Laboratory Quarterly Progress and Status Report* 30/2. Royal Institute of Technology, Stockholm. 1–80.
- Fletcher, Janet 2010. The Prosody of Speech: Timing and Rhythm. In Hardcastle, William J. – Laver, John – Gibbon, Fiona E. (eds.): *The handbook of phonetic sciences*. 2nd edition. Blackwell, Oxford. 521–602.
- Fox Tree, Jean E. – Schrock John C. 2002. Basic meanings of *you know* and *I mean*. *Journal of Pragmatics* 34. 727–747.
- Fraser, Bruce 1999. What are discourse markers? *Journal of Pragmatics* 31. 931–952.
- Glücksmanová, Helena. 2008. Spontaneous speech reconstruction. *WDS '08 Proceedings of Contributed Papers*. Part I. 134–138.
- Gósy Mária 2005. *Pszicholingvisztika*. Osiris Kiadó, Budapest.
- Gósy Mária 2009. Szóejtés és szóészlelés: változatosság és adaptálódás. *Beszédkutatás* 2009. 46–76.
- Gósy Mária 2013. Vannak-e reáliák a spontán beszédben? In Bárdosi Vilmos (szerk.): *Reáliák – A lexikológiától a frazeológiáig: Értelmezések és fordítási kérdések*. Tinta Kiadó, Budapest. 219–228.
- Gósy Mária – Horváth Viktória 2009. Hogyan tükrözi a kiejtés a nyelvi funkció változását? In Keszler Borbála – Tátrai Szilárd (szerk.): *Diskurzus a grammatikában – grammatika a diskurzusban*. Tinta Kiadó, Budapest. 37–45.
- Gósy Mária – Gyarmathy Dorottya – Horváth Viktória – Grácz Tekla Etelka – Beke András – Neuberger Tilda – Nikléczy Péter 2012. BEA: Beszélt nyelvi adatbázis. In Gósy Mária (szerk.): *Beszéd, adatbázis, kutatások*. Akadémiai Kiadó, Budapest. 9–25.
- Haader Lea 2001. Mikrodiakrónia és változásvizsgálat (összetett mondatokban). *Magyar Nyelvőr* 125/3. 354–371.
- Henton, Caroline – Bladon, Anthony 1988. Creak as a sociophonetic marker. In Hyman, Larry M. – Li, Charles N. (eds.): *Language, speech and mind. Studies in honour of Victoria A. Fromkin*. Routledge, London–New York. 3–29.
- Horváth Viktória 2009. *Funkció és kivitelezés a megakadásjelenségekben*. PhD-értekezés. ELTE, Budapest.
- Huszár Ágnes 1985. A rádió és a televízió beszélt nyelvének mondattana. In Grétsy László (szerk.): *Nyelvészet és tömegkommunikáció*. Tömegkommunikációs Kutatóközpont, Budapest. 73–117.
- Jucker, Andreas H. 1993. The discourse marker *well*: A relevance-theoretical account. *Journal of Pragmatics* 19. 435–452.
- Juhász Dezső 1992. A kötőszók. A módosítószók. In Benkő Loránd (szerk.): *A magyar nyelv történeti nyelvtana II/1*. Akadémiai Kiadó, Budapest. 772–838.

- Kassai Ilona 1979. *Időtartam és kvantitás a magyar nyelvben*. Nyelvtudományi Értekezések 112. Akadémiai Kiadó, Budapest.
- Kovács Magdolna 2002. *Tendenciák és szabályszerűségek a magánhangzó-időtartamok produkciójában és percepciójában*. Debreceni Egyetem Kossuth Egyetemi Kiadója, Debrecen.
- Lehiste, Ilse 1965. Juncture. In *Proceedings of the 5th International Congress of Phonetic Sciences, Münster 1964*. S. Karger, New York. 172–200.
- Lenk, Uta 1998. Discourse markers and global coherence in conversation. *Journal of Pragmatics* 30. 245–257.
- Magdics Klára 1966. A magyar beszédhangok időtartama. *Nyelvtudományi Közlemények* 68. 125–139.
- Markó Alexandra 2013. *Az irreguláris zöngé funkciói a magyar beszédben*. ELTE Eötvös Kiadó, Budapest.
- Markó Alexandra – Dér Csilla Ilona 2008. Magyar diskurzusjelölők korpuszalapú vizsgálata. In Bereczki András – Csepregi Márta – Klima László (szerk.): *Urálisztikai Tanulmányok 18. Ünnepi írások Havas Ferenc tiszteletére*. ELTE BTK Finnugor Tanszék – Numi-Tórem Finnugor Alapítvány, Budapest. 535–556.
- Pierrehumbert, Janet 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In Bybee, Joan L. – Hopper, Paul J. (eds.): *Frequency effects and the emergence of lexical structure*. Benjamins, Amsterdam. 137–157.
- Pusztai Ferenc szerk. 2003. *Magyar értelmező kéziszótár*. Akadémiai Kiadó, Budapest.
- Rácz Endre 1995. Nem mondatrészkifejező mellékmondatok. In Benkő Loránd (szerk.): *A magyar nyelv történeti nyelvtana II/2*. Akadémiai Kiadó, Budapest. 694–718.
- Schachter, Stanley – Christenfeld, Nicholas – Ravina, Bernard – Bilous, Frances 1991. Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology* 60. 362–367.
- Schourup, Lawrence 1999. Discourse markers. *Lingua* 107. 227–265.
- Slifka, Janet 2006. Some physiological correlates to regular and irregular phonation at the end of an utterance. *Journal of Voice* 20/2. 171–186.
- Szaszák György 2008. *A szupraszegmentális jellemzők szerepe és felhasználása a gépi beszédfelismerésben*. PhD-értekezés. BME, Budapest.
- White, Laurence – Mády, Katalin 2008. The long and the short and the final: Phonological vowel length and prosodic timing in Hungarian. In Barbosa, Plinio A. – Madureira, Sandra – Reis, César (eds.): *Proceedings 4th Speech Prosody Conference*. Campinas. 363–366.

A kutatást a 108762 számú OTKA-pályázat támogatta.

NONVERBÁLIS HANGJELENSÉGEK FIATALOK ÉS IDŐSEK SPONTÁN BESZÉDÉBEN

Bóna Judit

Bevezetés

Köztudott, hogy a beszéd a verbális tartalom mellett számos nem verbális vokális elemet is tartalmaz. Ezek megjelenhetnek nyelvi tartalommal együtt és nyelvi tartalom nélkül is (Vicsi et al. 2011). Kifejezhetjük velük érzelmeinket, hangulatunkat, véleményünket, vagy visszajelzést adhatunk általuk a beszédpartnerünknek; de vannak olyan nem verbális hangjelenségek is, amelyek nem bírnak jelentéssel (például szándéktalan testhangok, fiziológiai reflexek) (vö. Markó 2005, 2006; Vicsi et al. 2011; Neuberger 2012; Markó et al. 2014). Az érzelmi vagy egészségi állapotunkra utaló vokális jelzések, illetőleg az artikuláció természetes velejárójaként megjelenő testhangok szándéktalan elemei a beszédnek, míg bizonyos gesztusokat, hangokat (pl. véleményt kifejező nyelvcsettintést, rosszálló torokköszörülést), hűmmögéseket szándékosan hozunk létre (Vicsi et al. 2011; Neuberger 2012).

A nonverbális hangjelenségek vizsgálata a kommunikációban betöltött szerepén túl számos gyakorlati alkalmazás szempontjából igen fontos. Ezek a jelenségek ugyanis nehezíthetik a gépi beszédfelismerésben a felismerő algoritmusok pontos működését, míg a beszédszintézisben hozzájárulhatnak a természetes hangzás kialakításához (például Li et al. 2008; Prylipko et al. 2012; Neuberger–Beke 2013; Sárosi et al. 2014). Emellett a nonverbális vokális jelek tanulmányozása szerepet játszhat a beszélőfelismerésben, a beszélők megítélésében is (Mohammadi et al. 2010).

A nonverbális vokális jelzések gyakorisága, időtartama számos tényezőtől függ, ilyenek lehetnek például a beszélő személye, fizikai és érzelmi állapota, a beszédpartnerek ismeretsége vagy a beszéd típusa. Neuberger (2012) például igazolta, hogy a beszédhelyzet hatással van bizonyos hangjelenségek gyakoriságára: a narratívában a nyelvcsettintések, a társalgásban a nevetések előfordulása lett gyakoribb. A gyakoriság mellett a hangjelenségek időtartamát is befolyásolja a beszéd típusa (Neuberger 2012).

A jelen tanulmány célja, hogy megvizsgálja, hogyan változik egyes nonverbális vokális elemek megjelenése a beszédben a beszélők életkorától függően. Öt olyan nonverbális vokális jelenséget vizsgálunk, amelyek megjelenése – feltételezésünk szerint – függ(het) a beszélő életkorától. Ezek közül négy testhang, az artikuláció természetes velejárója: a hallható légzés, a köhögés, a nyelvcsettintés és a hallható nyelés.

Bár ennek a négy testhangnak a megjelenése számos tényezőtől, így például a beszélő egészségi állapotától is függhet (például egy náthás, influenzás megbetegedés esetén megváltozhat a légzés, illetve többet köhögnek a beszélők; vagy a nem megfelelő mennyiségű folyadékbevitel, illetve a fűtés is okozhatja a nyálkahártya kiszáradását, ezáltal a hangos nyelést, nyelvcsettintést, köhögést és torokköszörülést), mégis azt feltételezzük, hogy a természetes öregedés is hatással van a gyakoriságukra és az időtartamukra. Ugyanis a beszédszervek az életkor előrehaladtával számos változáson mennek keresztül (Balázs 1993). A tüdő kapacitása például csökken, ami gyakoribb levegővételhez és csökkent hangerőhöz vezethet (Levitzky 1984; Huber 2008). A lélegzetvételek – a szükséges levegőmennyiség biztosításán túl – utalhatnak bizonyos beszédtervezési folyamatokra, illetve jelezhetik a beszédszándékot, kijelölhetik a társalgási fordulók határait is (például McFarland 2001; Scobbie et al. 2011; Neuberger 2012; Rochet-Capellan–Fuchs 2013). Időskorban a hangszalagok is rugalmatlanabbá válnak (Balázs 1993), és ha beszéd közben kiszáradnak, gyakoribbá válik a torokköszörülés, köhögés (ezeknek a jelenségeknek természetesen lehet a közlésben más funkciójuk is, például kifejezhetnek rosszalást, vagy jelezhetik a beszélő zavarát, esetleg beszédtervezési nehézségét). A köhögéshez hasonló okai és funkciói lehetnek a hangos nyelésnek. A nyelvmozgást is befolyásolja az idősödés: az életkor előrehaladtával a nyelv működése lassabbá válik, ami az artikuláció pontatlanabbá válásához vezethet (Balázs 1993). A nyelvcsettintés egyrészt az artikuláció velejárója, másrészt funkciója lehet a szóátvétel, a közlés indításának a jelzése, illetve beszédtervezési nehézség feloldása (Neuberger 2012).

Az ötödik elemzett jelenség a beszélők érzelmeiről árulkodik, de megjelenését befolyásolja a téma, a beszélők személyisége és a beszédpartnerek ismeretisége, illetve a köztük lévő viszony is: ez a nevetés. A nevetésről szóló szakirodalom igen gazdag. Meghatározták például a típusait, az átlagos időtartamát, illetve az alaphangmagasságát, de vizsgálták az automatikus detektálásuk lehetőségeit is (Bickley–Hunnicut 1992; Provine 1993; Rothgänger et al. 1998; Bachorowski et al. 2001; Trouvain 2003; Neuberger–Beke 2013 stb.). A jelen vizsgálatban azért lehet szerepe az életkornak a nevetések előfordulásában, mert mind a fiatalokkal, mind az idősekkel fiatal kísérletvezető készített felvételt.

A jelen vizsgálatunk kérdései a következők: 1. Van-e különbség a nonverbális vokális elemek előfordulási gyakoriságában idősek és fiatalok beszéde között? 2. Vannak-e ezek között olyan elemek, amelyek inkább az idősekre vagy a fiatalokra jellemzőek? 3. Milyen időtartamban realizálódnak a különböző típusú jelenségek a két életkori csoportban?

Hipotéziseink szerint 1. az idősek beszédprodukcióiban gyakoribbak a nonverbális hangjelenségek, mint a fiatalok beszédében a beszédszervek említett változásai miatt. 2. Az elemzett jelenségek típusa befolyásolja a gyakoriságot: vannak olyanok, amelyek időseknél gyakoribbak, míg mások a fiatalok

loknál. 3. A nonverbális hangjelenségek időtartamai különbséget mutatnak a két életkori csoport között.

Anyag, módszer, kísérleti személyek

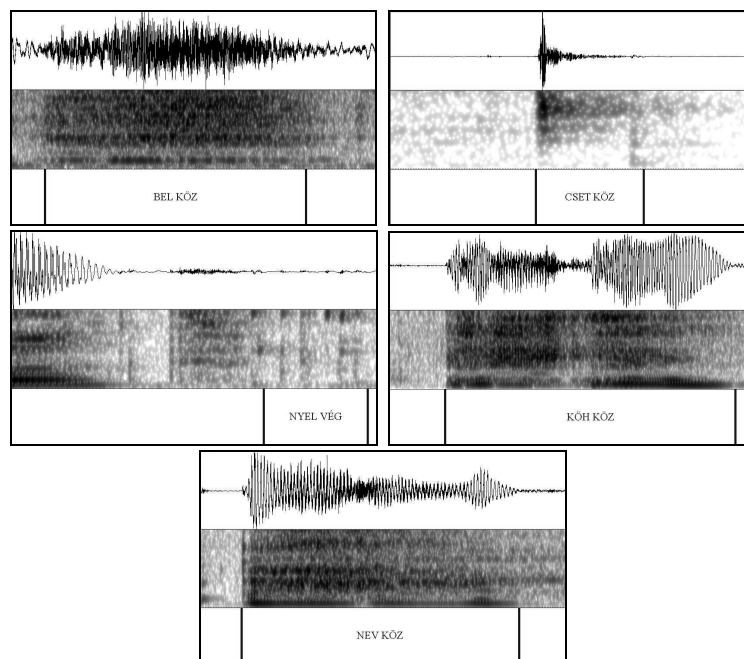
A kutatáshoz a BEA adatbázisból (Gösy et al. 2012) választottuk ki 10 fiatal (21–26 évesek, az átlagéletkoruk 23 év) és 10 idős (70–81 évesek, az átlagéletkoruk 75 év) hangfelvételét. Az adatközlők mindegyike ép hallású, magyar anyanyelvű volt; és mindkét életkori csoportban 5 nő és 5 férfi szerepelt. A kiválasztásukkor figyelembe vettük az iskolai végzettségüket is, minden adatközlő legalább érettségivel rendelkezett.

A BEA hangfelvételeiből a spontán beszédet választottuk ki elemzésre, minden adatközlőtől mintegy 5–5,5 percet. Összesen 104 percnyi hanganyagot elemeztünk, ez összesen 11 491 szónyi beszédanyag volt. A hangfelvételeken a beszélők a munkájukról, hobbijukról vagy a családjukról beszéltek; a kísérletvezető csak akkor szólalt meg, amikor az adatközlők nem tudták folytatni a beszédet.

A hangfelvételekben a Praat 5.0 szoftverrel (Boersma–Weenink 2008) annotáltuk a nonverbális jelenségeket; ezek a következők voltak: be- és kilélegzés, köhögés, nevetés, nyelvcsettintés vagy nyammogás (vö. Neuberger 2012) és nyelés. Az annotálásakor mind az auditív, mind a vizuális jeleket figyelembe vettük a hangjelenség kezdetének és lecsengésének jelöléséhez, mivel egy-egy nyelés vagy ki-belégzés határainak meghatározása csak az egyik csatorna alapján nem lett volna egyértelmű. Az 1. ábrán mindegyik jelenség annotálására szerepel egy példa. A be- és kilélegzés hallhatóságának megítélése viszonylag szubjektív. A jelen tanulmányban minden lélegzést jelző hangjelenséget felcímkeztünk, függetlenül attól, hogy milyen intenzitású volt (a kicsi intenzitású, de azért a hangszínképen látható, illetve halk hangerővel hallható jelenségeket is). Emiatt, illetve a beszélők közötti egyéni különbségek miatt lehetséges az, hogy a jelen anyagban nagyobb számban fordulnak elő a levegővételre utaló nonverbális jelzések, mint például Neuberger (2012) tanulmányában.

A nevetések megjelenhetnek két beszédszakasz között, de verbális elemekkel együtt is. Ezért ezen jelenségek időtartamának meghatározásakor a nevetős beszédből azt a szakaszt mértük meg, amelyet a beszélő nevetve mondott ki. Amint azt korábban már leírtuk, a nevetések előfordulását befolyásolhatja a beszélgetőpartnerek közötti viszony. A jelen vizsgálatban figyeltünk arra, hogy olyan felvételeket válasszuk ki elemzésre, amelyekben az adatközlő és a felvételvezető nem ismerték egymást – legalábbis a megszólításokból, a kérdésekből erre lehetett következtetni.

Összesen 1415 jelenséget elemeztünk. Ebből 570 a fiataloknál, 845 az időseknél fordult elő. Voltak olyanok, amelyek minden adatközlőnél megjelentek (például a hangos lélegzés), míg más típusok csak egyes beszédprodukciókban (1. táblázat).



1. ábra

A vizsgált jelenségek annotálása (BEL = belégzés, CSET = nyelvcsettintés, NYEL = nyelés, KÖH = köhögés, NEV = nevetés; KÖZ = beszédszakaszok között, VÉG = beszédszakasz végén)

1. táblázat: Az elemzett jelenségek száma, illetve az adatközlők száma és aránya, akik produkálták a jelenséget

	Fiatalok		Idősek	
	Darab	Beszélők száma (fő)	Darab	Beszélők száma (fő)
Lélegzés	393	10 (100%)	671	10 (100%)
Köhögés	6	2 (20%)	13	5 (50%)
Nyelvcsettintés	121	10 (100%)	143	10 (100%)
Nyelés	28	7 (70%)	15	3 (30%)
Nevetés	22	7 (70%)	2	2 (20%)

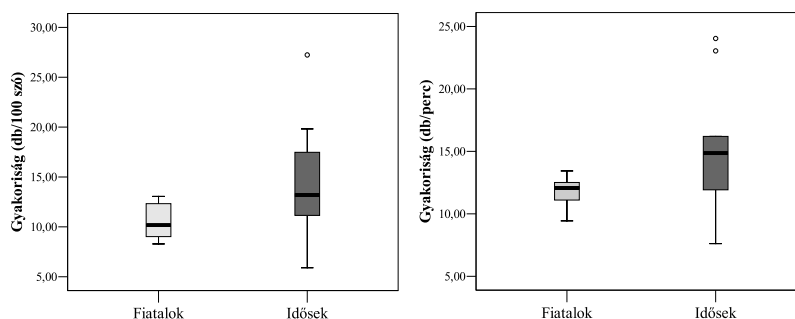
Elemeztük ezen jelenségek gyakoriságát (100 szóra és percre vetítve is), időtartamát, illetve megjelenési helyét. A gyakoriság kétféle kiszámítási módja több szempontból is indokolt volt. Egyrészt az idősek és a fiatalok szignifikánsan eltérő beszéd- és artikulációs tempója miatt (vö. Bóna 2013) a percre vetített gyakoriság és a szószámra vetített gyakoriság értékeiben lehetnek elté-

rések. Előfordulhat például az, hogy egy beszélő ritkábban vesz levegőt egy perc alatt, mint egy másik, ugyanakkor a lassabb artikulációs tempójának köszönhetően kevesebb szót ejt ki két lélegzetvétel között, ami miatt a 100 szóra vetített belégzéseinek a száma nagyobb lesz. Másrészt azért is indokolt a kétféle érték megadása, mert a szakirodalomban mindkét kiszámítási módot szokás alkalmazni. A más szakirodalmi adatokkal való összehasonlíthatóság is indokolta a jelen tanulmányban mindkét gyakorisági érték megadását.

Az adatokon statisztikai elemzést (eloszlástól függően egytényezős és ismételt méréses ANOVA-t, Kruskal–Wallis-tesztet, Mann–Whitney *U*-tesztet, illetve Friedman-tesztet és Wilcoxon-tesztet) is végeztünk az SPSS 13.0 szoftverrel 95%-os konfidenciaszinten.

Eredmények

Az összes elemzett jelenség gyakorisága a 2. ábrán olvasható. A fiataloknál átlagosan 10,4 nonverbális hangjelenség fordult elő 100 szóban, ez átlagosan 11,7 megjelenést jelentett percenként. Az időseknél 100 szóra vetítve 14,6 jelenséget adatoltunk, ami percenként 15,2-es gyakoriságnak felelt meg. Az idős életkori csoportnál tehát gyakoribbak voltak a nonverbális vokális jelenségek, mint a fiataloknál, és ezt a statisztikai próba is megerősítette [egytényezős ANOVA 100 szóra vetítve a gyakoriságot: $F(1, 19) = 4,729$; $p = 0,043$; percire vetítve a gyakoriságot: $F(1, 19) = 4,480$; $p = 0,048$].



2. ábra

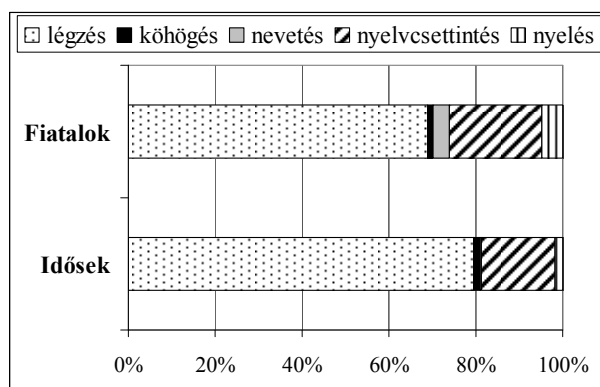
Az összes nonverbális hangjelenség gyakorisága életkoronként
100 szóra vetítve (balra) és percire vetítve (jobbra)

Az összes jelenségen belül mindkét életkori csoportnál a be- és kilégzés volt a leggyakoribb, míg a legkisebb arányban a fiataloknál a köhögés és a nevetés, az időseknél pedig a nevetés fordult elő (3. ábra).

A be- és kilégzés

Mind a belégzés, mind a kilégzés együtt járhat hallható és a hangszínképen látható hanggal. A lélegzésre utaló hangjelenség mindegyik adatközlő beszéd-

produkciójában megjelent. A kétféle irányú légzés közül a belégzés gyakrabban volt hallható, mint a kilégzés, mindkét életkori csoportban. A fiataloknál a belégzés 88,0%-át tette ki az összes hallható lélegzésnek, míg az időseknél a belégzések aránya 88,4% volt. A hallható lélegzés gyakoribb volt az időseknél, mint a fiataloknál (2. táblázat), amit a statisztikai próba is alátámasztott [egytényezős ANOVA 100 szóra vetítve a gyakoriságot: $F(1, 19) = 6,262$; $p = 0,022$; percra vetítve a gyakoriságot: $F(1, 19) = 6,621$; $p = 0,019$].



3. ábra

A különböző típusú jelenségek aránya a két életkori csoport beszédprodukcióiban

2. táblázat: A hallható lélegzések gyakorisága életkori csoportonként

Életkori csoport	Lélegzés/100 szó		Lélegzés/perc	
	Átlag	Min–max	Átlag	Min–max
Fiatalok	7,2	5,3–10,0	8,0	5,8–10,0
Idősek	11,7	3,5–22,6	12,1	4,6–19,9

A belégzések döntő többsége (a fiataloknál 93,4%, az időseknél 88,2%) a beszédszakaszok között fordult elő, kisebb százalékuk jelent meg a beszédszakaszok végén (a fiataloknál 5,2%, az időseknél 9,8%), és csak igen kis arányuk közvetlenül a beszédszakaszok elején (a fiataloknál 1,4%, az időseknél 2,0%). A kilégzések többsége a beszédszakaszok legvégén jelent meg (a fiataloknál 74,5%, az időseknél 91,0%), kisebb arányban fordultak elő két beszédszakasz között (a fiataloknál 19,1%, az időseknél 7,7%), és csak kivételes esetben adatoltuk őket beszédszakaszok elején (a fiataloknál 3 esetben, ami a kilégzések 6,4%-a; az időseknél 1 esetben, ami a kilégzések 1,3%-a). A levegővételek elsősorban a beszédhez szükséges levegőmennyiséget, illetve a szöveg tagolását biztosították.

Elemeztük a lélegzetvételek időtartamát is. Az összes lélegzetvétel időtartama szignifikáns különbséget mutatott a két életkori csoport között (Mann–Whitney U -teszt: $Z = -6,724$; $p < 0,001$); a fiataloknál hosszabb átlagos időtartamot (átlag: 423 ms; szórás: 192 ms) mértünk, mint az időseknél (átlag: 355 ms; szórás: 183 ms).

Külön megvizsgáltuk a belégzések és a kilégzések időtartamát is (3. táblázat). Csak a belégzések időtartamát elemezve is szignifikáns különbség volt a fiatalok és az idősek között (Mann–Whitney U -teszt: $Z = -5,717$; $p < 0,001$), a fiatalok szignifikánsan hosszabb ideig lélegeztek be, mint az idősek. A belégzés helye a beszéd folyamatban nem befolyásolta szignifikánsan az időtartamot egyik életkorban sem, de megfigyelhető, hogy a legrövidebb átlagos időtartam a beszédszakaszok elején megjelenő belégzésekre volt jellemző mindkét életkori csoportban.

3. táblázat: A belégzések és a kilégzések időtartama (ms)

	Fiatal		Idős	
	Átlag	Min–max	Átlag	Min–max
Belégzés				
Az összes jelenség	419	160–1650	366	40–1510
Beszédszakasz elején	349	230–630	300	40–560
Beszédszakaszok között	420	160–1650	369	120–1510
Beszédszakasz végén	434	200–850	350	130–740
Kilégzés				
Az összes jelenség	448	100–1730	271	60–1530
Beszédszakasz elején	360	230–490	1063	–
Beszédszakaszok között	737	240–1730	493	320–660
Beszédszakasz végén	382	100–1040	241	60–1530

A kilégzések időtartamában is szignifikáns különbség volt az életkori csoportok között (Mann–Whitney U -teszt: $Z = -4,465$; $p < 0,001$), a fiatalokra hosszabb átlagos kilégzési idő volt jellemző. A kilégzés helye szignifikáns hatással volt az időtartamra: a két beszédszakasz között mért kilégzések (mintegy sóhajtások) szignifikánsan hosszabb tartamúak voltak, mint a beszédszakaszok végén mérték. A fiataloknál: Mann–Whitney U -teszt: $Z = -2,895$; $p = 0,004$; az időseknél: Mann–Whitney U -teszt: $Z = -3,174$; $p = 0,002$ (a kis számú adat miatt a beszédszakaszok elején adatolt kilégzéseket nem lehetett figyelembe venni a statisztikai elemzés során).

A köhögés

Köhögést igen kis számban adatoltunk. A fiataloknál mindössze két adatközlőnél jelent meg, beszélőnként 3-3, összesen 6 előfordulással, míg az idősek csoportjában öt beszélőnél adatoltuk a jelenséget, összesen 13 előfordulással. Ez az összes adatközlő beszédprodukciónak figyelembe véve a fiatalok

loknál 0,1 db/100 szó (0–0,7 db/100 szó) és 0,1 db/perc (0–0,6 db/perc); az időseknél 0,2 db/100 szó (0–1,2 db/100 szó) és 0,2 db/perc (0–1,4 db/perc) gyakoriságot jelentett. A kis elemszám miatt statisztikai próbát nem végeztünk. A köhögések többsége két beszédszakasz között jelent meg (a fiataloknál 83,3%, az időseknél 69,2%), az összes többi esetben pedig a beszédszakaszok elején adatoltuk őket. Bár a beszélők szándékosan is létrehozhatnak köhögést (például a zavaruk feloldására vagy rosszallásuk kifejezésére), a vizsgált hanganyagban csak fiziológias okok miatt jelent meg ez a jelenség.

A fiatalok köhögése tendenciaszerűen hosszabb időtartamban valósult meg, mint az időseké (4. táblázat), szignifikáns különbség nem volt az időtartamokban a két életkori csoport között.

4. táblázat: A köhögések időtartama (ms)

Életkori csoport	Időtartam (ms)	
	Átlag	Min–max
Fiatalok	556	206–816
Idősek	442	172–716

A nyelvcsettintés

A hallható légzés után a második leggyakoribb jelenség a nyelvcsettintés volt, kivétel nélkül minden adatközlőnél megjelent. Az idősek átlagosan gyakrabban produkálták ezt a jelenséget, mint a fiatalok (5. táblázat). A nagy egyéni különbségek miatt azonban nem volt szignifikáns különbség a két életkori csoport között. A fiataloknál a nyelvcsettintések többsége (73,6%) a beszédszakaszok elején, 26,4%-a pedig két beszédszakasz között jelent meg. Az időseknél a leggyakoribb előfordulás (60,1%) két beszédszakasz között volt; a nyelvcsettintések 39,2%-a pedig a beszédszakaszok elején fordult elő. Az időseknél adatoltunk egy olyan esetet (0,7%) is, amikor a beszédszakasz végén jelent meg a csettintés.

5. táblázat: A nyelvcsettintések gyakorisága életkori csoportonként

Életkori csoport	db/100 szó		db/perc	
	Átlag	Min–max	Átlag	Min–max
Fiatalok	2,2	1,3–3,2	2,5	1,8–4,2
Idősek	2,4	0,3–5,5	2,6	0,3–5,7

A nyelvcsettintések többféle funkciót töltöttek be a beszédprodukciókban. A beszélő például a mentális lexikonban való keresésre használta a hezitálással együtt a következő megnyilatkozásban: *érdekes öm □ [nyelvcsettintés] élmény volt mert...* Egy másik beszélőnek a nyelvi tartalom megtervezéséhez volt szüksége hosszabb időre (összesen 3464 ms-ra). Ez idő alatt hangos nyelés, nyelvcsettintés és hezitálás is megjelent a néma szünetek mellett a beszé-

dében: *ez tulajdonképpen egy ilyen* [nyelés] □ [nyelvcsettintés] □ *öö pihenési időszak volt*. Előfordult olyan is, hogy a beszélő a közlés indításakor produkált nyelvcssettintést.

A fiatalok nyelvcssettintései tendenciaszerűen hosszabb időtartamban valósultak meg, mint az időseké (6. táblázat), szignifikáns különbség azonban nem volt az időtartamokban a két életkori csoport között. A különböző fonetikai pozíciókban megjelenő nyelvcssettintések időtartamában a fiataloknál nem volt szignifikáns a különbség; míg az idők esetében a két beszédszakasz között megjelenő nyelvcssettintések időtartama szignifikánsan hosszabb volt, mint a beszédszakaszok elején állóké (Mann–Whitney U -teszt: $Z = -3,690$; $p < 0,001$).

6. táblázat: A nyelvcssettintések időtartama (ms)

	Fiatal		Idős	
	Átlag	Min–max	Átlag	Min–max
Az összes jelenség	107	20–460	94	20–350
Beszédszakasz elején	105	20–460	79	20–350
Beszédszakaszok között	111	20–250	102	30–200
Beszédszakasz végén	–	–	219	–

A nyelés

Hallható nyelés csak az adatközlők felénél fordult elő. A fiataloknál hét fő beszédében, 28 alkalommal; az időseknél három beszélőnél, 15-ször adatoltuk. Ez az összes beszélőt és felvételt figyelembe véve a fiataloknál 0,5 db/100 szó (0–2,1 db/100 szó) és 0,6 db/perc (0–2,1 db/perc); az időseknél 0,3 db/100 szó (0–1,2 db/100 szó) és 0,3 db/perc (0–1,3 db/perc) gyakoriságot jelentett. Nem volt statisztikai különbség a két csoport között a gyakoriságban. Ahogyan a nyelvcssettintés egyik példájában már megfigyelhető volt, hallható nyelés létrejöhetett időnyerési stratégiaként a beszélő tervezési bizonytalanságának feloldására, de megjelenhetett egyszerű fiziológiás szükségletként is.

A hallható nyelések döntő többsége két beszédszakasz között jelent meg (a fiataloknál 89,3%, az időseknél 100%), egy esetben (3,6%) a fiataloknál adatoltuk beszédszakasz elején, illetve kétszer (7,1%, szintén a fiataloknál) közvetlenül a beszédszakasz végén. A nyelések időtartama (7. táblázat) a fiataloknál szignifikánsan hosszabb volt, mint az időseknél (Mann–Whitney U -teszt: $Z = -2,599$; $p = 0,009$).

7. táblázat: A hallható nyelések időtartama (ms)

Életkori csoport	Időtartam (ms)	
	Átlag	Min–max
Fiatalok	232	70–401
Idősek	175	46–549

A nevetés

A nevetés funkcióját tekintve elkülönül a többi nonverbális vokális elemtől, hiszen egyértelműen érzelemkifejező funkciójú. A megjelenését befolyásol(hat)ják a beszélők tulajdonságai (menyire közvetlen, vidám), a beszédpartnernek közötti viszony, illetve a beszéd témája is. Ezért is lehetséges az, hogy bár mindkét életkori csoportban ritkán adatoltuk ezt a jelenséget, a fiataloknál sokkal gyakrabban jelent meg, mint az időseknél. A BEA adatbázis felvételeit ugyanis egy fiatal felvételvezető készíti, akivel könnyebben, gyorsabban egymásra hangolódtak a fiatal adatközlők, és így többet nevettek együtt. Mivel a vizsgált beszéd típus narratíva volt, így csak a beszélő saját közléséhez kapcsolódó nevetéseket adatoltunk a hanganyagban.

Nevetés a fiataloknál hét adatközlőnél jelent meg, összesen 22 előfordulással. Ezzel szemben az idősek csoportjában mindössze két beszélőnél adatoltuk a jelenséget, összesen 2 előfordulással. Ez az összes adatközlő beszédprodukciját figyelembe véve a fiataloknál 0,35 db/100 szó (0–0,9 db/100 szó) és 0,45 db/perc (0–1,2 db/perc); az időseknél 0,04 db/100 szó (0–0,2 db/100 szó) és 0,04 db/perc (0–0,2 db/perc) gyakoriságot jelentett. A statisztikai elemzés szerint szignifikáns különbség volt a gyakoriságban a két életkori csoport között (Mann–Whitney *U*-teszt: $Z = -2,895$; $p = 0,004$).

A nevetés nemcsak önmagában, hanem a verbális közlés közben is megjelenhet, ekkor a beszélő nevetve közli mondanivalóját. A fiatalok nevetéseinek 45%-a volt ilyen, míg az időseknél 50% (a két adatból egy). A fiatalok nevetéseinek átlaga 622 ms (290–1852 ms), az idősek nevetéseinek időtartama 329 ms és 2148 ms volt (az utóbbi beszéd közbeni érték).

Az egyes hangjelenségek összevetése azonos életkori csoporton belül

Végezetül összevetettük az azonos életkori csoportokon belül a különböző típusú jelenségek gyakoriságát és időtartamát. Mind a fiataloknál (Friedman-próba: 100 szóra vetítve $\chi^2 = 34,619$; $p < 0,001$; percre vetítve $\chi^2 = 34,619$; $p < 0,001$), mind az időseknél (Friedman-próba: 100 szóra vetítve $\chi^2 = 35,055$; $p < 0,001$; percre vetítve $\chi^2 = 38,515$; $p < 0,001$) szignifikáns különbség volt a gyakoriságban az egyes típusok között. Egy-egy típus gyakoriságát összevetve (8. táblázat) a Wilcoxon-próba mindkét életkori csoportban szignifikáns különbséget mutatott a légzés és a köhögés; a légzés és a nyelvcsettintés; a légzés és a nyelés, a nyelvcsettintés és a nyelés; illetve a nyelvcsettintés és a köhögés között. A nevetés gyakorisága a fiatalok esetében a légzéstől (itt a normál eloszlás miatt ismételt méréses ANOVA-t használtunk), a köhögéstől, illetve a nyelvcsettintéstől különbözött szignifikánsan. Az időseknél a nevetés gyakorisága csak a légzéstől, illetve a nyelvcsettintéstől mutatott statisztikai eltérést.

Az időtartamok összevetésekor csak a lélegzést és a nyelvcsettintést tudtuk figyelembe venni, mivel csak ez a két típus jelent meg minden beszélő beszédprodukcijában. Mivel ismétléses próbát használtunk, beszélőnként kiszámítottuk mind a lélegzések, mind a nyelvcsettintések beszélőre jellemző átlagos időtartamát. A statisztikai próba szerint mind a fiataloknál (Wilco-

xon-próba: $Z = -2,803$; $p = 0,005$), mind az időseknél [ismételt mérések ANOVA: $F(1, 9) = 521,680$; $p < 0,001$] szignifikáns volt a különbség a két típus átlagidőtartamai között.

8. táblázat: A statisztikai próba eredményei (gyakoriság)

	Fiatalok		Idősek	
	db/100 szó	db/perc	db/100 szó	db/perc
Lélegzés–	$Z = -2,803$	$Z = -2,803$	$Z = -2,803$	$Z = -2,934$
köhögés	$p = 0,005$	$p = 0,005$	$p = 0,005$	$p = 0,003$
Lélegzés–	$Z = -2,803$	$Z = -2,803$	$Z = -2,701$	$Z = -2,845$
nyelvcsettintés	$p = 0,005$	$p = 0,005$	$p = 0,007$	$p = 0,004$
Lélegzés–nyelés	$Z = -2,803$	$Z = -2,803$	$Z = -2,803$	$Z = -2,934$
	$p = 0,005$	$p = 0,005$	$p = 0,005$	$p = 0,003$
Lélegzés–	$F(1, 9) = 272,2$	$Z = -2,803$	$Z = -2,803$	$Z = -2,934$
nevetés	$p < 0,001$	$p = 0,005$	$p = 0,005$	$p = 0,003$
Köhögés–	$Z = -2,803$	$Z = -2,803$	$Z = -2,803$	$Z = -2,934$
nyelvcsettintés	$p = 0,005$	$p = 0,005$	$p = 0,005$	$p = 0,003$
Köhögés–nyelés	–	$Z = -2,028$ $p = 0,043$	–	–
Köhögés–	–	–	–	$Z = -2,023$ $p = 0,043$
nevetés	–	–	–	–
Nyelvcsettintés–	$Z = -2,803$	$Z = -2,803$	$Z = -2,803$	$Z = -2,934$
nyelés	$p = 0,005$	$p = 0,005$	$p = 0,005$	$p = 0,003$
Nyelvcsettintés–	$Z = -2,803$	$Z = -2,803$	$Z = -2,803$	$Z = -2,934$
nevetés	$p = 0,005$	$p = 0,005$	$p = 0,005$	$p = 0,003$
Nyelés–nevetés	–	–	–	–

Az időtartamok átlagait beszélőktől függetlenül figyelembe véve mindkét életkori csoportban felállítottunk egy sorrendet, amely azt mutatja, hogyan viszonyul egymáshoz az eltérő típusú nonverbális hangjelenségek tartama. (Bár az átlagok sokszor félrevezetőek lehetnek a szélsőértékek miatt, esetünkben mégis jelzik a vizsgált jelenségekre jellemző tendenciákat.) A két életkori csoportban azonos sorrend alakult ki azzal a különbséggel, hogy az időseknél a nevetést nem vettük bele az elemzésbe. A sorrend tehát a következő: *nyelvcsettintés* < *nyelés* < *légzés* < *köhögés* < *nevetés*. Adataink megerősítik Neuberger (2012) fiatalok spontán narratíváira és társalgásaira kapott eredményeit.

Következtetések

Kutatásunkban nonverbális hangjelenségek gyakoriságát és időtartamát elemeztük az életkor függvényében. Az első hipotézisünk, amely az összes elemzett jelenség gyakoriságára vonatkozott, igazolódott: az idősek beszédében gyakrabban fordultak elő a vizsgált jelenségek. Ugyanakkor a nonverbális

hangjelenségek típusa meghatározta azok előfordulását, és az egyes típusok gyakorisága eltérő volt a két életkorban (második hipotézis). Az időseknél szignifikánsan gyakoribb volt a hangos lélegzés, mint a fiataloknál, és tendenciaszerűen gyakrabban fordult elő a köhögés és a nyelvcsettintés. Ez a korábban említett fiziológiai, biológiai változásokra vezethető vissza, azaz a csökkenő tüdőkapacitásra, a szárazabb nyálkahártyákra, a beszédszervek renyhébb működésére (Balázs 1993). A fiataloknál szignifikánsan gyakoribb volt a nevetések előfordulása, ami a beszédpartnerek azonos életkorára és így a felszabadultabb társalgásra vezethető vissza (vö. Bata–Grácsi 2008). A hangos nyelés előfordulása is tendenciaszerűen gyakoribb volt a fiataloknál; az egyes esetek elemzése azt mutatja, hogy ezeket a beszédtervezési bizonytalanságok feloldására használták az adatközlők. A korábbi kutatások pedig azt mutatják, hogy a fiatalok beszédében gyakoribbak a beszédtervezési bizonytalanságra utaló elemek (megakadásjelenségek), mint az időseknél (Bóna 2014).

Az egyes jelenségek időtartamában is találtunk különbséget a két életkori csoport között (harmadik hipotézis). Csak a testhangokat figyelembe véve (a nevetést a kis elemszám miatt ebből a szempontból nem vizsgáltuk) azt találtuk, hogy a fiatalok mindegyik jelenséget hosszabb időtartamban valósították meg, mint az idősek. Szignifikáns különbség volt a lélegzés és a nyelés időtartamában a két életkori csoport között, és csak tendenciaszerű a köhögés és a nyelvcsettintés esetében. Ezek az adatok is megerősítik például azt, hogy az idősek tüdőkapacitása kisebb (Balázs 1993), rövidebb ideig lélegeznek be, és emiatt gyakrabban kell levegőt venniük.

A kutatásunk bár csak tíz-tíz adatközlő beszédét veti össze, mégis fontos tanulságul szolgálhat a gyakorlat számára. A nonverbális hangjelenségek további elemzése ugyanis hozzájárulhat például a beszélői profilalkotáshoz, a beszélő életkorának pontosabb felismeréséhez a beszéd alapján.

Irodalom

- Bachorowski, Jo-Anne – Smoski, Moria J. – Owren, Michael J. 2001. The acoustic features of human laughter. *Journal of the Acoustical Society of America* 110/3. 1581–1597.
- Balázs Boglárka 1993. Az időskori hangképzés jellemzői. *Beszédkutatás* '93. 156–165.
- Bata Sarolta – Grácsi Tekla Etelka 2008. A beszédpartner életkorának hatása a beszéd szupraszegmentális jellegzetességeire. In Keszler Borbála – Tátrai Szilárd (szerk.): *Diskurzus a grammatikában – grammatika a diskurzusban*. Tinta Kiadó, Budapest. 74–83.
- Bickley, Corine – Hunnicut, Sheri 1992. Acoustic analysis of laughter. In *Proceedings of the International Conference on Spoken Language Processing*. Banff, Canada. 927–930.
- Boersma, Paul – Weenink, David 2008. Praat: doing phonetics by computer (Version 5.0.1) http://www.fon.hum.uva.nl/praat/download_win.html (A letöltés ideje 2008. május 5.)

- Bóna Judit 2013. *A spontán beszéd sajátosságai az időskorban*. ELTE Eötvös Kiadó, Budapest.
- Bóna Judit 2014. Megakadásjelenségek az életkor, a nem és a beszéd típus függvényében. *Beszéd kutatás* 2014. 123–143.
- Gósy Mária – Gyarmathy Dorottya – Horváth Viktória – Gráci Tekla Etelka – Beke András – Neuberger Tilda – Nikléczy Péter 2012. BEA: Beszélt nyelvi adatbázis. In Gósy Mária (szerk.): *Beszéd, adatbázis, kutatások*. Akadémiai Kiadó, Budapest. 9–24.
- Huber, Jessica E. 2008. Effects of utterance length and vocal loudness on speech breathing in older adults. *Respiratory physiology & neurobiology* 164/3. 323–330.
- Johar, Swati 2014. Paralinguistic profiling using speech recognition. *International Journal of Speech Technology* 17/3. 205–209.
- Levitzky, Michael G. 1984. Effects of aging on the respiratory system. *Physiologist* 27/2. 102–107.
- Li, Yanxiong – He, Qianhua – Li, Tao – Wang, Weining 2008. A detection method of lip-smack in spontaneous speech. In: *International Conference on Audio, Language and Image Processing, ICALIP 2008*. IEEE. 292–297.
- Markó Alexandra 2005. „Szavak nélkül”. Nonverbális vokális közlések fonetikai elemzése. *Magyar Nyelvőr* 129. 88–104.
- Markó Alexandra 2006. Nonverbális vokális jelek a társalgásban. *Beszéd kutatás* 2006. 57–68.
- Markó, Alexandra – Gósy, Mária – Neuberger, Tilda 2014. Prosody patterns of feedback expressions in Hungarian spontaneous speech. In: *Social and Linguistic Speech Prosody: Proceedings of the 7th international conference on Speech Prosody*. Science Foundation Ireland, Dublin. 482–486.
- McFarland, David H. 2001. Respiratory markers of conversational interaction. *Journal of Speech, Language, and Hearing Research* 44/1. 128–143.
- Mohammadi, Gelareh – Vinciarelli, Alessandro – Mortillaro, Marcello 2010. The voice of personality: Mapping nonverbal vocal behavior into trait attributions. In: *Proceedings of the 2nd international workshop on Social signal processing*. ACM. 17–20.
- Neuberger Tilda 2012. Nonverbális hangjelenségek a spontán beszédben. In Gósy Mária (szerk.): *Beszéd, adatbázis, kutatások*. Akadémiai Kiadó, Budapest. 215–235.
- Neuberger, Tilda – Beke, András 2013. Automatic laughter detection in spontaneous speech using GMM–SVM method. In: *Text, Speech, and Dialogue 2013*. Springer, Berlin–Heidelberg. 113–120.
- Provine, Robert R. 1993. Laughter. *American Scientist* 84. 38–45.
- Prylipko, Dmytro – Vlasenko, Bogdan – Stolcke, Andreas – Wendemuth, Andreas 2012. Language modeling of nonverbal vocalizations in spontaneous speech. In: *Text, Speech and Dialogue 2012*. Springer, Berlin–Heidelberg. 488–495.
- Rochet-Capellan, Amélie – Fuchs, Susanne 2013. The interplay of linguistic structure and breathing in German spontaneous speech. In: *Proceedings of Interspeech*. Paper 1228.
- Rothgänger, Hartmut – Hauser, Gertrud – Cappellini, Aldo Carlo – Guidotti, Assunta 1998. Analysis of laughter and speech sounds in Italian and German students. *Naturwissenschaften* 85. 394–402.

- Sárosi, Gellért – Tarján, Balázs – Fegyő, Tibor – Mihajlik, Péter 2014. Automated transcription of conversational Call Center speech – with respect to non-verbal acoustic events. *Intelligent Decision Technologies* 8/4. 265–275.
- Scobbie, James M. – Schaeffler, Sonja – Mennen, Ineke 2011. Audible aspects of speech preparation. In: *Proceedings of 17th ICPhS, Hong Kong*. 1782–1785.
- Trouvain, Jürgen 2003. Segmenting phonetic units in laughter. In: *Proceedings of the International Congress of Phonetic Sciences (ICPhS)*, Barcelona, Spain. 2793–2796.
- Vicsi Klára – Sztahó Dávid – Kiss Gábor 2011. Nem verbális hangjelenségek spontán társalgásban. *Beszéd kutatás 2011*. 134–147.

A tanulmány a Bolyai János Kutatási Ösztöndíj támogatásával készült.

A METAPRAGMATIKAI TUDATOSSÁG JELZÉSEI SZÁMÍTÓGÉP KÖZVETÍTETTE TÁRSALGÁSI NARRATÍVÁKBAN

Laczkó Krisztina – Tátrai Szilárd

Bevezetés

A tanulmány az internet közvetítette spontán írásbeliségben megjelenő történetelbeszélésekben vizsgálja a metapragmatikai tudatosság megjelenését és ennek jellemző mintázatait. Első lépésként a vizsgálat elméleti háttér-feltevéseit explikáljuk. Társas kognitív nyelvészeti kiindulópontból a narratív diskurzusokat mint közös figyelmi jeleneteket, a történeteket pedig mint referenciális jeleneteket mutatjuk be, ennek kontextusában értelmezve a metapragmatikai tudatosság, vagyis a nyelvi tevékenységhez való reflexív viszonyulás jelenségét. Ezt követően korpuszvizsgálat alapján mutatjuk be a számítógép közvetítette társalgási narratív diskurzusok egyik legjellegzetesebb műfajspecifikus jellemzőjét: a történetmondás szituáltságának a reflektáltságát, azaz a komplex narratív nyelvi tevékenységre történő reflexiók működését. A korpusz és a vizsgálati szempontok bemutatását követően ismertetjük az empirikus vizsgálat eredményeit. A tanulmányban a legtipikusabb nyelvi megvalósulásokra irányítjuk rá a figyelmet: egyrészt a metapragmatikai tudatosság jelzéseinek jellegzetes típusaira, másrészt a metapragmatikai jelzések szemantikai kidolgozottságára, harmadrészt a jelzéstípusoknak a narratívákban belüli elhelyezkedésére fókuszálunk. A tanulmányt a legfontosabb következtetések, tanulságok levonása zárja.

Elméleti háttér

A funkcionális kognitív nyelvészet egymással termékeny diskurzust folytató nyelvreírési modelleket foglal magában, amelyek közös jellemzője, hogy a nyelvtant nem autonóm rendszerként értelmezik, hanem a beszédtevékenység felől közelítik meg, lényegi szerepet tulajdonítva az abban szerepet kapó kognitív folyamatoknak (l. pl. Croft–Cruse 2004; Geeraerts–Cuycens eds. 2007; Evans–Poursel eds. 2009; Kövecses–Benczes 2010; Tolcsvai Nagy 2013). A nyelvi rendszer elemeinek működését így a mindenkori beszélő nézőpontjából, természetes diszkurzív közegükből kiindulva, a jelentésképzésben betöltött szerepüket előtérbe helyezve mutatják be (Tolcsvai Nagy 2010: 11–13). Ugyanakkor ezek a modellek kérdésfeltevéseikben és közelítésmódjaikban részben el is térnek egymástól. Az utóbbi időszakban több elméleti munka – elsősorban bizonyos társas-lélektani, illetve pragmatikai

belátások hatására – a nyelv funkcionális kognitív leírásában a hangsúlyt a nyelvi megismerés társas (interakciós és interszubjektív) alapjaira, illetve annak következményeire helyezi (l. pl. Sinha 1999, 2009; Tomasello 2002, 2003, 2011; Croft 2009). Ez a társas kognitív nyelvészeti kiindulópont termékenyen összehangolható azzal a pragmatikai nézőponttal, amely a nyelvhasználat kognitív és szociokulturális feltételeit együttesen, egymásra vonatkoztatva láttatja, a nyelvhasználatot társas megismerő tevékenységként írja le, és a kontextusfüggő, dinamikus jelentésképzés problémáját helyezi a középpontba (l. Verschueren 1999; Sandra et al. eds. 2009; Tátrai 2011, 2013).

A kontextusfüggő, dinamikus jelentésképzés folyamatának a közös figyelmi jelenetként megvalósuló diskurzusok biztosítanak keretet (l. Tomasello 2002; Sinha 2005). A diskurzusokat ugyanis – legyenek azok akár mindennapi társalgások, akár irodalmi szövegek – általánosságban jellemzi, hogy résztvevők valamely természetes nyelv (vagy nyelvek) közegében interakcióba lépnek egymással, és a másik figyelmének irányításával, illetve követésével együttesen hozzák létre azokat a referenciális jeleneteket, amelyekkel a kommunikációs igényeik kielégítését célozzák meg (vö. Verschueren 1999; Verschueren–Brisard 2009). A közös figyelmi jelenet tehát a szimbolikus nyelvi kommunikáció alapvető feltétele, az abban való részvétel olyan interszubjektív emberi tevékenység, amelyben lehetővé válik a másik figyelmének ráirányítása a világ dolgaira és eseményeire (a referenciális jelenetre) és ezáltal mentális irányultságának (figyelmének, megértésének) befolyásolása (bővebben l. Tátrai 2011: 29–35). A nyelvi szimbólumok így a világgal kapcsolatos tapasztalatok fogalmi konstruálásának különböző lehetőségeit teremtik meg (l. Langacker 2008: 55–89; Verhagen 2007). A nyelvi szimbólumok egyfelől leképezik a fogalmilag különféleképpen megkonstruált tapasztalatokat, másfelől kezdeményezik is a tapasztalatok különböző fogalmi konstruálásait (l. Sinha 1999). Természetesen nincs ez másként a történetek elmesélésekor sem.

Számos olyan narratív diskurzus létezik, amelyek meghatározó jellegzetessége, hogy bennük, illetőleg általuk a világgal kapcsolatos tapasztalatainkat úgy konstruáljuk meg fogalmilag, és tesszük mások számára hozzáférhetővé nyelvileg, hogy azok történetekké szerveződnek. Habár a nyelvi megismerést tekintve is fontos tanulsága van annak, hogy a tapasztalatok narratívaként történő fogalmi konstruálására más médiumok is lehetőséget adnak (l. Herman 2009), itt a figyelemirányítás kérdése a szimbolikus nyelvi kommunikáció narratív diskurzusaival összefüggésben merül fel. A történetek megértése sajátos értelmezői viszonyulást követel meg a befogadótól (hallgatótól, olvasótól), mivel a nyelvi szimbólumokból kiindulva olyan – jellemzően perceptuálisan nem megfigyelhető – referenciális jelenetet kell konceptuálisan feldolgoznia, amelyben a szereplőket térben és időben nyomon tudja követni, a cselekvéseik és a velük történetek között időbeli és oksági összefüggéseket tud találni (Brown 1994: 15–28). A narratív diskurzus során alkalma-

zott nyelvi szimbólumok referenciális értelmezése, vagyis a közös figyelmi jelenet interszubjektív kontextusában történő episztemikus lehorgonyzása ennél fogva szorosan összefügg a diskurzusvilágban történő tájékozódás lehetőségeivel, a történetmondó perspektívájának a működésével. A narratív diskurzusokban történetként megjelenő referenciális jelenet megértésére ugyanis alapvető befolyással van az, hogy a diskurzus résztvevői közül ki és hogyan irányítja a tapasztalatok fogalmi megkonstruálását, a referenciális jelenet feldolgozását, azaz ki és hogyan jelöli ki, hogy honnan reprezentálódnak az elbeszélte események. Eszerint az alapvető kérdés, hogy honnan és hogyan konstruálódik a történet szereplőinek konceptuálisan feldolgozható fizikai és társas világa, amelyben a szereplők cselekvései és a velük történetek megvalósulnak, valamint a mentális világuk, amelyben a szereplők aktív tudati folyamatai zajlanak (l. Tátrai 2011: 171–189; vö. még Bruner 1986: 14). E tanulmányban azonban a központi kérdésünk nem az, hogy mi és hogyan irányítja a figyelmiünket akkor, amikor megkíséreljük megérteni a különböző történeteket, hanem az, hogy mindez hogyan válhat metapragmatikai reflexió tárgyává.

A metapragmatikai tudatosság fogalma a diskurzus résztvevőinek reflexív viszonyulását jelöli a nyelvi tevékenységhez, illetve az annak közegében zajló dinamikus jelentésképzéshez (bővebben l. Tátrai 2011: 119–125). Az éppen megnyilatkozó személyek ugyanis reflexió tárgyává tehetik a saját, a másik, illetőleg az egyéb harmadik személyek megnyilatkozásait, illetve befogadói tevékenységét, továbbá reflektálhatnak magára az éppen folyó diskurzusra, illetve annak szerveződésére is. Mindez azt jelenti, hogy a résztvevők képesek reflexíven viszonyulni a különféle nyelvi reprezentációkhoz és a velük összefüggő társas kognitív folyamatokhoz, illetőleg szociokulturális elvárásokhoz. A metapragmatikai tudatosságnak vannak megfigyelhető nyelvi nyomai, amelyeket metapragmatikai jelzéseknek nevezünk. A metapragmatikai jelzések szemantikai kidolgozottsága ikonikus összefüggést mutat a tudatosság mértékével, hiszen a metapragmatikai reflexiók nagyobb mértékű szemantikai kidolgozottsága a megnyilatkozó nagyobb mértékű metapragmatikai tudatosságáról tanúskodik. A metapragmatikai tudatosság ugyanakkor nem egyszerűen nyelvi jelzések alkalmazását jelenti, hanem a résztvevők különböző mértékű reflexív viszonyulását a közös nyelvi tevékenységükhöz, a dinamikus jelentésképzéshez, amelyre e jelzésekkel a megnyilatkozó felhívja a figyelmet.

Anyag és módszer

A vizsgálat korpuszát egy, az internet közvetítette műfaj szolgáltatta, az úgynevezett tematikus topik, jelen esetben egy, a történetelbeszélést középontba állító diskurzus. A tér-idő és a résztvevői szerepviszonyokat illetően fontos megjegyezni, hogy a megnyilatkozók és a befogadók soha nincsenek ugyanabban a fizikai térben, és a sikeres diskurzushoz nincsen szükség arra sem, hogy ugyanabban az időben legyenek, hiszen a megnyilatkozások archi-

válódnak, visszakereshetők és újra feldolgozhatók. A megnyilatkozók anonim módon jelennek meg a társalgásban, többnyire nem ismerik egymást, így alapvetően a nyelvi megnyilatkozásaik révén ítélik meg egymást. Annak, hogy ezt a korpuszt választottuk a vizsgálathoz, az alapvető oka az volt, hogy a magyar nyelvre vonatkozóan nem áll rendelkezésünkre olyan spontán beszélt nyelvi anyag, amelyben a történetmondás bármilyen funkcionális kognitív pragmatikai aspektusból vizsgálható lenne. Azok a szövegek azonban, amelyek az internet közvetítette spontán írásbeliséget (l. Petykó 2012) képviselik, például a chat, a fórumszövegek vagy a kommentek, megfelelő anyagszámot biztosítanak. (A *spontán írott nyelv* terminust Petykó Márton [2012] javasolta a magyar szakirodalomban korábban használt *netnyelv*, *másodlagos írásbeliség* vagy *írott beszélt nyelv* terminusok helyett.)

Az általunk korpuszként használt szöveg címe: Beégésem története. Az egyik legnagyobb magyar hírportál, az Index fórumán található, kizárólag magyar nyelvű megnyilatkozásokat tartalmaz (<http://forum.index.hu/Article/showArticle?t=9017476&la=125481821>). A topik elindítója a következőképpen tematizálja célját az első megnyilatkozásban:

- (1) Széleskörű tapasztalat, hogy személytelenül az ember sokkal könnyebben beszél számára kellemetlen dolgokról is, sokkal hajlamosabb az öniróniára. Nos, ez a témakör igazolhatja a fenti véleményt. Arra kérek ugyanis minden ide látogatót, osszátok meg a többiekkel éleetek legnagyobb leégéseinek történetét, amikor tényleg ott álltatok megszeppenülve, és mindenki rajtatok röhögött. Minél nagyobb az égés, annál jobb a sztori!

A diskurzus 2000 májusában indult el, és a tanulmány írásának idején 26 095 hozzászólást tartalmaz. A megnyilatkozások 83,6%-a a tematizálásnak megfelelően történetelbeszélés, ezt egészítik ki diszkurzív módon a történetekre reagáló hozzászólások, megnyilatkozások, esetenként a témától eltérő társalgások. A fenti diskurzusból 200 hozzászólás 228 narratív történetelbeszélését választottuk ki a vizsgálathoz, ezeket időrendi sorrendet követve számozással külön fájlban rögzítettük. A történetek átlagosan 230 szövegszót tartalmaznak, a legrövidebb 36, a leghosszabb 425 szóból áll. A megállapításaink szándékosan a legjellegzetesebb kvalitatív jellemzőket veszik sorra, egy árnyaltabb, részletező, kvantitatív szempontokat is érvényesítő vizsgálat megalapozásaként. Az elemzés során módszertanilag úgy jártunk el, hogy három, egymással összefüggő szempontból vizsgáltuk meg a narratív nyelvi tevékenységre történő metapragmatikai reflexiókat. (A metapragmatikai reflexiók vizsgálata a korábbi magyar nyelvű szakirodalomban szépirodalmi narratívák vonatkozásában történt meg, l. Tátrai 2002.) Az elemzési szempontok a következők voltak: a) Milyen típusú metapragmatikai jelzések fordulnak elő a vizsgált narratív megnyilatkozásokban? b) Mi mondható el a korpuszról a metapragmatikai

reflexiók szemantikai kidolgozottságának tekintetében?
 c) A történetmondás melyik pontján, az elején, közben vagy a végén helyezkednek-e el a metapragmatikai jelzések egyes típusai?

A három szempont nem egymástól elkülönülten, hanem sajátos összefüggérendszerként reprezentálható, amely ily módon összetett mintázatok felvázolására is lehetőséget ad az internet közvetítette történetelbeszélés vonatkozásában.

Eredmények

A vizsgált narratívák általános jellegzetessége, hogy magára a beszédese-ményre számos, műfajspecifikus utalás történik. Az, hogy a történetmondó nyelvi tevékenysége metapragmatikai reflexió tárgyává válik, kiemelt, feltűnő jellegzetessége az internet közvetítette történetelbeszélésnek. Egyértelműen megállapítható, hogy a metapragmatikai tudatosság mértéke nagyfokú, a diskurzus erőteljesen kiaknázza a nyelvi tevékenységre történő reflektálás lehetőségét: a 228 történetelbeszélésből mindössze 16 nem tartalmaz a történetmondás aktusára reflektáló metapragmatikai jelzést, azaz 212 esetben találunk szemantikailag különböző mértékben kidolgozott reflexiókat a narratív szituációra. Az alábbiakban a különböző reflexiótípusokra fókuszálunk részletesebben, és ezek alapján mutatjuk be a metapragmatikai tudatosság mintázatát az internetes történetelbeszélésben, a szemantikai kidolgozottság szerepére is kitérve.

A metapragmatikai tudatosság jelzéseiinek jellegzetes típusai

1. A történetmondói tevékenységre történő reflexió, amelynek legnyilvánvalóbb jelölői az egyes szám első személyű mondat jelentő igék és az ezekkel konstruált, személydeixissel lehorgonyzott szerkezetek (ehhez l. Tátrai 2011: 131–135; Laczkó–Tátrai 2012: 236–243) (a példákat eredeti helyesírással közöljük):

- (2) Hali, én is mondok egyet. Nem annyira égő, inkább vicces.
- (3) Ígéretemhez híven megosztom pár beégésem veletek.
- (4) Mindenkinek, illetve neked Gro, „megtérésed” örömeinek okán egy saját történettel hozakodnék elő, remélem nem fogok csalódást okozni.
- (5) Hu, csak azért írok, mert az előbbi óvszeres sztorin enyhébb röhögőgörcsöt kaptam...
- (6) Na nyomok még egyet. Van bőven.
- (7) Na akkor en is elohozakodok eggyel. Eleg kemény beeges volt, raadásul nem is az en hibamból.

A történetmondó profiláltan ekkor saját nyelvi tevékenységére reflektál, nyelvi tevékenységének előtérbe kerülését a mondást jelentő igék reprezentálják: *elmondok még egyet; megpróbálom szavakkal visszaadni; megosztom pár beégésem; írogatok ide történetet; saját történettel hozakodnék elő; idemásolom a sztorit; előhozakodok eggyel; riogatlak titeket a történetemmel; remélem, még nem mondtam el nektek; régen szóltam hozzá, gazdagítom a fórumot.*

2. A történetbefogadói tevékenységre történő reflexió, amelynek jelzései a korpuszban jellemzően elsősorban a többes szám második személyű ugyan-csak személydeixissel lehorgonyzott formák:

- (8) kb. 10-12 évvel ezelőtt cirkusz jött a faluba, ahol akkor laktam. tudjátok, az a csóró fajta cirkusz, aminek fő attrakciója egy darab rühes, fogatlan teve szokott lenni
- (9) Ti akartátok: Balatonlelle, 2000 nyara
- (10) El tudjátok képzelni a szituációt, amikor megálltam a piros lámpánál, a körülöttem lévő autók megdöbbent, vagy inkább lenéző tekintettel bámultak, mivel úgy tűnt, hogy valami állat nyomja veszettül a gázt, hogy mindenkit letoljon az útról
- (11) Tudnotok kell, hogy idegeneket néha még akkor sem engedtünk be, ha többen voltunk, nemhogy csak én egyedül.
- (12) Na akkor ezt figyeljétek: Kanadában jártunk még úgy 6 éve.
- (13) Mostmár nem is tunik annyira jónak, de higgyétek el, akkor baromi ciki volt.

Ebben az esetben jellegzetesen a befogadó mentális tevékenységére kognitív igékkel reflektál a megnyilatkozó: tudjátok; ti akartátok; na ezt figyeljétek; na akkor kapaszkodjatok; talán emlékeztek; ezt add össze; hát ez fogjátok meg.

3. A történetmondás mint esemény is válhat reflexió tárgyává, akár a történetmondói nyelvi tevékenység, akár a befogadói kognitív tevékenység kiemelésével:

- (14) ezt nehéz lesz szavakban visszaadni
- (15) Azt még tudni kell, hogy ekkor már a két lány két éve tanult németül a suliban
- (16) mondani se kell, hogy a hölgyemény pont az első szótagnál kezdett odafigyelni arra, hogy mi folyik a telefonvonalban
- (17) kár, hogy itt nem lehet érzékeltetni a hangsúlyt

Ilyenkor a történetmondót, illetve a befogadót nem a személy-, hanem az idődeixis művelete horgonyozza le az adott szituációban a jelen idejű igealakok használatával (l. Tátrai 2011: 139–142). E metapragmatikai reflexiók a mondást jelentő vagy kognitív igeneveket tartalmazó szerkezetekként konstruálódnak meg a megnyilatkozásban (*lehet, hogy ezt nem kellene elmesélni; remélem, érthető lesz*).

4. Sajátos metapragmatikai tudatosságot képvisel az a helyzet, amikor a történetmondói reflexió arra vonatkozik, hogy a beszélő nem saját maga által átélt eseményt mesél el. Ez tekinthető az idézés olyan formájának, amelyben a történet újramondására és egyben másvalakinek a narratív tevékenységére reflektál a megnyilatkozó:

- (18) Egy ismerősöm mesélte ezt:
- (19) haverok mesélték a következő sztorit, szerintem nem igaz (őket ismerve), de nem is ez a lényeg
- (20) Nagybátyám beégése, máséval tündökölni könnyebb, nem?
- (21) En meg itt (is) új vagyok tehát... nem az en sztorim nem rólam fog szólni!

Ennek anyagunkban relatív nagy számát és az újramondás jelzésének erős előtérbe helyezését az is indokolja, hogy az internetes topik a megnyilatkozók saját tapasztalatainak a megosztását kéri, ahogy ezt már a címevel is tematizálja. Példák még az idézésre reflektáló metapragmatikai jelzésekre, azaz a történetmondás történetére: *tesóm mesélte egyszer; a következő történet egy buszvezető ismerőssémmel esett meg; kolléganőm mesélte, hogy mi történt a férjével meg a közös fiúkkal*.

5. A saját vagy másik (korábbi, esetleg későbbi) narratívára történő reflexiók diskurzusdeixisek alkalmazásával válnak lehetővé (l. Tátrai 2011: 142–144; Laczkó–Tátrai 2012: 248–250), amelyek vonatkozhatnak magára a teljes történetre, de annak egyes részeire is. Az ilyen reflexió jellemzően az *ez* mutató névmással valósul meg. Az egyik jellemző eset a névmás önálló használata:

- (22) ez ma történt
- (23) erről eszembe jutott egy másik történet
- (24) Ezt nehéz lesz szavakban visszaadni, de megpróbálom...
- (25) Ebből is látszik, hogy régi történet.
- (26) Ez később még fontos lesz.

A másik lehetőség, amikor a névmási elem összekapcsolódik egy már határozott névelővel lehorgonyozódó főnévvel, amely magát a történetet konstruálja meg fogalmilag valamilyen formában: *ez a történet (sztori, eset, ízelítő, epizód, beégés)*. Minthogy a határozott névelő lehorgonyozza a főnevet a szerkezetben, a névmás diskurzusdeixisként funkcionáló metapragmatikai jelzéssé válik, amely a következőkben elmesélt történetre reflektál kataforikusan (ritkábban anaforikusan), elsődleges szerepe a diskurzus szerveződésére történő reflektálás. A referenciális jelenet egészére vagy részére reflektáló szerkezet ugyanakkor elliptikus is lehet.

(27) Nem égés, de jó!

(28) Logikus, nem?

(29) Friss termés, de nem én égtem.

Ezekben a ritkábban előforduló esetekben a közelre mutató névmás nem vesz részt a metapragmatikai reflexió megkonstruálásában. A diskurzusdeixis funkciójú elemeknek természetesen csak egyik legjellegzetesebb megvalósulása a mutató névmásnak ez a használata. Néhány további példa:

(30) A barátom mesélte *következő sztorit*

(31) oké, legyenek katonasztorik, *az első*:

6. A metapragmatikai jelzések közé tartoznak még a különféle diskurzusjelölők (*nos, na, namost, szóval, hát, ugye*), amelyek ugyancsak magára a diskurzusnak a szerveződésére reflektálnak, szekvenciákra osztják, és egyfajta kapcsolófunkciójuk van a diskurzusrészek között, így fontos szerepet töltenek be a figyelemirányításban:

(32) Szóval par hónappal ezelőtt Perthben (Ausztrália) tanultam és akkoriban probalgattam angolul beszélni... mar amennyire.

(33) Nos, ott vagyunk, nem rohanunk, videozgattunk kicsit, stb.

(34) Na, akkora már eléggé dögglődtünk, de még volt pénzmag.

(35) Akkor leesett a dolog, sűrű bocsánatkérés... Hát égett a fejem.

Ezek a metapragmatikai jelzések, amelyekre a vizsgált narratív megnyilatkozások is számos példát szolgáltatnak, többnyire erősen deszemantizáltak.

A metapragmatikai jelzések szemantikai kidolgozottsága

Elemzésünk másik szempontja a típusok vizsgálata mellett a szemantikai kidolgozottságra vonatkozik, elsődlegesen a kidolgozottság fokára. A fokozatosság skalaritást feltételez, és alapvetően megállapítható, hogy a személy-

deixisként vagy szituatív lehoronyzottan megvalósuló metapragmatikai reflexiók képviselik a szemantikailag leginkább kidolgozott pólust (ilyenek például az egyes szám első személyű mondat jelentő igét vagy a második személyű kognitív igét tartalmazó megnyilatkozások; 1. 1. és 2. típus), a legkevésbé kidolgozottak pedig értelem szerűen a deszemantizálódott semantikus diskurzusjelölők (1. 6. típus). Köztes fokozatot mutatnak a személydeixissel le nem horgonyzott történetmondói tevékenységre reflektáló megvalósulások, valamint a mutató névmási diskurzusdeixis. Ebben a kérdéskörben azonban a fenti általánosítható megállapítások mellett még szükség lesz további árnyalásra. Most nézzünk meg három jellegzetes példát:

- (36) Itt az ideje, hogy én is gazdagítsam ezt a fórumot egy bejegyzéssel, ami nem az enyém, ezt a sztorit a barátom mesélte el nekem.
- (37) Logikus, nem?
- (38) A doktornő a legkisebb betegségre is vagy háromféle gyógyszert felírt. Nomármost, az én influenzámra is kaptunk egy csomót.

Az (36) példában a beszélő történetmondói tevékenysége a személyes névmás és az igei személyrag folytán lehoronyzódik, a *gazdagítsam a fórumot* szerkezet explikálja a történetmondás tevékenységét. Ezzel a történetmondó objektíválja saját nyelvi tevékenységét benne saját magával, a referenciális jelenet részévé teszi. Így a referenciális jelenet nemcsak a történetből (a megfigyelt jelenetből) áll, hanem egy metaszintet (a megfigyelői jelenetet) is magában foglal, amely az elbeszélte eseményeket megjelenítő elemi mondatoktól elkülönülő elemi mondatként reprezentálódik (l. Tátrai 2011: 121; vö. Langacker 2002). Mindemellett a közös figyelemi jelenetből a történet maga is objektíválódik a *bejegyzés* és a *sztori* megjelöléssel, a beszélő továbbá egy diskurzusdeixissel explikálttá teszi a beszédcselekmény terét (*ezt a fórumot*), és reflektál arra, hogy a történet újramondása történik, a történet kitől ered, és hangsúlyossá is teszi, egy tagadó formával személydeixissel lehoronyozva, hogy nem a saját tapasztalata (*nem az enyém*). Ha csak a metapragmatikai tudatosság reflexióit vesszük, nemcsak magas szemantikai kidolgozottsági fokozatról beszélhetünk, hanem a reflexió komplex is, hiszen a történetmondói tevékenység mellett a történet idézettségére is reflektál.

A (37) példa ugyancsak komplex ebben a vonatkozásban, hiszen reflektál magára a referenciális jelenetre (elliptikusan), valamint a befogadó mentális megértői tevékenységére (*logikus*). Mindezt azonban személydeixissel nem horgonyozza le, pusztán a jelen idő lehoronyzó szerepe jelzi a metapragmatikai funkciót (a történetek jellemzően a múltban horgonyzódnak le). Így szemantikailag kevésbé kidolgozott megvalósulásnak tekinthető, de a mondat a megfigyelői jelenetet itt is objektíválja.

A (38) példa egyértelmű diskurzusjelölő elemet tartalmaz, sem a kidolgozottság foka (deszemantizált elem), sem komplexitása nem nagyfokú, szerepe a szekvenciaváltás a doktornő jellemvonásának ábrázolásáról az események menetére.

A metapragmatikai tudatosság jelzései elhelyezkedése

Végezetül azt tekintjük át röviden, hogy a metapragmatikai reflexiók elhelyezkedése a narratívákhoz képest milyen jellegzetes mintázatot mutat az internetes történetelbeszélésben.

A szöveg kezdetén található a legtöbb metapragmatikai jelzés, amelyek szemantikai kidolgozottságukat tekintve a többi helyzethez képest, mind komplexitásban, mind fokozatosságukban a leginkább kidolgozottak, a történetmondó deiktikusan lehorgonyozódik, és a nyelvi tevékenység a mondat jelentő igékben is meglehetősen gyakran kifejeződik. Ugyancsak gyakori emellett a diskurzusdeiktikus utalás a történet egészére az *ez* névmással vagy ennek kifejtő formájában, a történet szemantikailag kidolgozottabb objektívizációjával (*ez a történet*). A korpuszban csak itt található a történet idézettségére vonatkozó metapragmatikai reflexió. A történetmondó nyelvi tevékenységénél sokkal kisebb számban történik reflexió a befogadói tevékenységre, amelyek megjelennek, azok viszont többnyire szemantikailag kidolgozottan, lehorgonyozva. Példák:

- (39) Ground, csak azért, mert itt ölég sok szegedi van, ide írom be a sztorit.
- (40) Ez még általánosban történt, az egy óras úttörőünnepségek idején.
- (41) A kolléganom mesélte, mi történt a férjével, meg a közös fiúkkal.
- (42) talán emlékeztek a régi szép időkre (4-5 éve), mikor is a városban az éjszakai mulatozások éjfélkor azzal értek véget, hogy kidobták az embert a bezárni vágyó szeszharapóból

Ezzel szemben a történet zárásaként már sokkal ritkábban jelennek meg a metapragmatikai jelzések, a mennyiségi különbség jelentős, körülbelül egyharmaddal kevesebb, mint a kezdő helyzetben. Ekkor vagy az addigi beszédtevékenységre történik visszautalás, vagy a következő lehetséges beszédtevékenységre előreutalás, vagy a narratív megnyilatkozás zárása explicálódik elsősorban. Az első esetben a visszautalás jellegzetesen szemantikailag kifejtetten a megelőző történetmondói tevékenységre vagy diskurzusdeixissel a referenciális jelenet egészére reflektál, azt téve explicitté, hogy a történetmondó megfelelt a műfaji követelményeknek (*ez nekem elég égés volt*), kivételesen ritka a befogadói mentális tevékenységére történő reflexió vagy a metapragmatikai idézés. Jellegzetes záró névmási diskurzusdeixissel történik annak a jelzése, hogy vége a referenciális jelenetnek vagy a közös figyelmi jelenet is lezárult (*ennyi*).

- (43) Na ennyit errol. MAjd ha meg eszembe jut mondok. Mer szivni szivtam eleget :)
- (44) Szóval... ennyi... a következőt talán majd legközelebb.
- (45) Hát ezt fogjátok meg! Durva, mi?
- (46) Mindezt Megyerinétől, a kolesz kezdetektől fogva fungáló igazgató-nőjétől hallottam.

Jelentős mennyiségű a metapragmatikai jelzés szöveg közben, ezek azonban körülbelül 80%-ban sematikus jelentésszerkezetű diskurzusjelölők. A más típusú metapragmatikai reflexiók közül nagyobb számban találunk még névmási diskurzusdeiktikus elemeket a referenciális jelenet egyes részeire reflektálva, a történetmondói tevékenység reflexiója azonban nem jelenik meg, a befogadói tevékenységre reflektálás viszont kis számban előfordul:

- (47) Rögtön hívja a hölgyet, akihez jött, erre kitalálhatjátok, ki lépett ki az ajtón.

Az elrendeződési mintázat gyakoriságának vélhetően az az oka, hogy habár a topik erőteljes tematizáltsága és a számítógép képernyőjén az írott megjelenésű szöveg elrendeződése kevésbé kívánná meg a történetelbeszélés kezdetének és végének metapragmatikai jelzését, a műfaj hagyományozódó sémái, a figyelemirányítás jelölésének beszélői igénye azonban különösen a kezdet jelölésében a narratív szituáció reflektáltságát eredményezi.

Következtetések

A vizsgálat során olyan – a funkcionális kognitív pragmatika háttérfeltevéseit érvényesítő – modellt vettük alapul, amely a történetként megkonstruált referenciális jelenet szerveződését a közös figyelmi jelenet interszubjektív kontextusához való viszonyából kiindulva értelmezi. Az internet közvetítette spontán írásbeliségben megjelenő történetelbeszélésben korpuszmintán elemeztük a metapragmatikai tudatosság megjelenését, amelyet elsősorban a reflexió típusa szerint részletezve és a szemantikai kidolgozottság fokozatossága alapján vizsgáltunk, majd a történetben elfoglalt helyzetük szerinti mintázatot igyekeztünk kimutatni. Eszerint a történetelbeszélői nyelvi tevékenységre és a befogadói mentális tevékenységre történő reflexiók a legkidolgozottabbak és nagy gyakorisággal a történet elejét jellemzik, a történet elmesélése közben a diskurzusjelölők dominálnak, a záró részben ismét visszatérnek a történetelbeszélői reflexiók, ám jóval kisebb számban, szemantikailag kevésbé kidolgozottan. A névmási diskurzusjelölők mindegyik részben megtalálhatók, de jelentősen nagyobb a számuk a kezdő részben, mint a záróban, ahol a referenciális jelenet egészére reflektálnak, szemben a közbülső helyzettel, ahol pedig kizárólag a referenciális jelenet részeire történik reflexió. A fenti

megállapítások alapvetően műfajspecifikusak, ám a mintázatok, azok elhelyezkedése, szemantikai kidolgozottságuk mértéke vélhetően követi a spontán beszélt nyelvi társalgások során kialakult sémákat, azaz ezek a sémák olyan szövegműfajokban is továbbhagyományozódnak, amelyek kevésbé igényelnék meg a megjelenésüket. Mivel azonban erre vonatkozó pontos kutatások még nem történtek, az összevetés, a hasonlóságok és a különbségek korpuszalapú vizsgálata a kutatás következő lépése kell, hogy legyen.

Irodalom

- Brown, Gillian 1994. Modes of understanding. In Brown, Gillian – Malmkjaer, Kirsten – Pollitt, Alastair – Williams, John (eds.): *Language and Understanding*. Oxford University Press, Oxford. 10–20.
- Bruner, Jerome 1986. *Actual minds, possible words*. Harvard University Press, Cambridge MA.
- Croft, William 2009. Towards a social cognitive linguistics. In Evans, Vyvyan – Poursel, Stephanie (eds.): *New directions in cognitive linguistics*. John Benjamins, Amsterdam. 395–420.
- Croft, William – Cruse, Alan D. 2004. *Cognitive linguistics*. Cambridge University Press, Cambridge.
- Evans, Vyvyan – Poursel, Stephanie eds. 2009. *New directions in cognitive linguistics*. John Benjamins, Amsterdam.
- Geeraerts, Dirk – Cuyckens, Hubert eds. 2007. *The Oxford handbook of cognitive linguistics*. Oxford University Press, Oxford.
- Herman, David 2009. Beyond voice and vision: cognitive grammar and focalization theory. In Hühn, Peter (ed.): *Point of view, perspective, and focalization: Modeling mediation in narrative*. De Gruyter, Berlin. 119–142.
- Kövecses Zoltán – Benczes Réka 2010. *Kognitív nyelvészet*. Akadémiai Kiadó, Budapest.
- Laczkó Krisztina – Tátrai Szilárd 2012. Személyek és/vagy dolgok. A harmadik személyű és a mutató névmási deixis a magyarban. In Tolcsvai Nagy Gábor – Tátrai Szilárd (szerk.): *Konstrukció és jelentés*. ELTE, Budapest. 231–257.
- Langacker, Ronald W. 2002. Deixis and subjectivity. In Brisard, Frank (ed.): *Grounding. The epistemic footing of deixis and reference*. Mouton, Berlin–New York. 1–28.
- Langacker, Ronald W. 2008. *Cognitive grammar. A basic introduction*. Oxford University Press, Oxford.
- Petykó Márton 2012. Az írott beszélt nyelvtől a spontán írott nyelv felé. In Hattyár Helga – Hugyeczi Enikő – Krepsz Valéria – Vladár Zsuzsa (szerk.): *A sokszínű alkalmazott nyelvészet. Tanulmányok az alkalmazott nyelvészet területeiről*. Tinta Könyvkiadó, Budapest. 44–52.
- Sandra, Dominik – Östman, Jan-Ola – Verschueren, Jef eds. 2009. *Cognition and pragmatics. Handbook of pragmatics highlights 3*. John Benjamins, Amsterdam, Philadelphia.

- Sinha, Chris 1999. Grounding, mapping and acts of meaning. In Janssen, Theo – Redeker, Gisela (eds.): *Cognitive linguistics: foundations, scope and methodology*. Mouton, de Gruyter. Berlin, New York. 223–255.
- Sinha, Chris 2005. Biology, culture and the emergence and elaboration of symbolization. In Saleemi, Anjum P. – Bohn, Ocke-Schwen – Gjedde, Albert (eds.): *Search of a language for the mind-brain: Can the multiple perspective unified?* Aarhus University Press, Aarhus. 311–335.
- Sinha, Chris 2009. Language as a biocultural niche and social institution. In Evans, Vyvyan – Pourcel, Stéphanie (eds.): *New directions in cognitive linguistics*. John Benjamins, Amsterdam, Philadelphia. 289–310.
- Tátrai Szilárd 2002. *Az „Én” az elbeszélésben – A perszonális narráció szövegtani megközelítése*. Argumentum Kiadó, Budapest.
- Tátrai Szilárd 2011. *Bevezetés a pragmatikába. Funkcionális kognitív megközelítés*. Tinta Kiadó, Budapest.
- Tátrai Szilárd 2013. Funkcionális pragmatika és kognitív nyelvészet. *Magyar Nyelv* 109/2. 197–204.
- Tolcsvai Nagy Gábor 2010. *Kognitív szemantika*. Konstantin filozófus Egyetem, Nyitra.
- Tolcsvai Nagy Gábor 2013. *Bevezetés a kognitív nyelvészetbe*. Akadémiai Kiadó, Budapest.
- Tomasello, Michael 2002. *Gondolkodás és kultúra*. Osiris Kiadó, Budapest.
- Tomasello, Michael 2003. *Constructing a language. A usage based theory of language acquisition*. Harvard University Press, Cambridge MA.
- Tomasello, Michael 2011. *Mi haszna az együttműködésnek?* Gondolat Kiadó, Budapest.
- Verhagen, Arie 2007. Construal and perspectivization. In Geeraerts, Dirk – Cuyckens, Hubert (eds.): *The Oxford handbook of cognitive linguistics*. Oxford University Press, Oxford. 48–81.
- Verschueren, Jef 1999. *Understanding pragmatics*. Arnold, London–New York–Sydney–Auckland.
- Verschueren, Jef – Brisard, Frank 2009. Adaptability. In Verschueren, Jef – Östman, Jan-Ola (eds.): *Key notions for pragmatics. Handbook of pragmatics highlights 1*. John Benjamins, Amsterdam, Philadelphia. 28–47.

KISISKOLÁSOK ALAPHANGMAGASSÁGÁNAK VARIABILITÁSA

Beke András – Horváth Viktória

Bevezetés

A beszédkutatás egyik legtöbbet vizsgált területe kezdetek óta az alaphangmagasság jellemzőinek és változásainak leírása. A magyar beszédre vonatkozó kutatások még a közelmúltban is elsősorban felolvasott szövegen és mondatokon alapultak (vö. pl. Bolla 1992; Gósy–Terkén 1994; Varga 1994, 2002; Olasz 2002). A kutatások egy része beszédtechnológiai felhasználással készült, a gépi beszéd előállításának egyik fő célja ugyanis a prozódia megfelelő megvalósítása (pl. Olasz 1995, 2002). A beszédfelismerésben szintén fontos szerepet játszik a prozódia (Szászák 2008). A technikai fejlődésnek köszönhetően az utóbbi két évtizedben megindult a spontánbeszéd-korpuszok kiépítése, így lehetőség nyílt a spontán beszéd alaphangszerkezetének leírására. A gazdagréti kábeltelevízió válogatott adásairól Varga László készített intonációs átiratot (1988). Markó Alexandra saját korpuszán elemezte többek között az alaphangmagasság jellemzőit a szövegtípus függvényében (2005).

A BEA spontánbeszéd-adatbázis (Gósy et al. 2012) lehetővé tette az alaphangszerkezet vizsgálatát jó minőségű, nagy mennyiségű anyagon. Beke (2008a, 2008b) a felolvasás és a spontán beszéd alaphangszerkezetét hasonlított össze, illetve a beszélőfelismeréshez modellezte az alapfrekvencia-eloszlást. Markó (2009) vizsgálata szerint a stigmatizált szökö dallamzár gyakrabban fordult elő spontán beszédben, mint felolvasásban. A fiatal és idős nők spontán beszédének és felolvasásának elemzése azt mutatta, hogy a beszédmód nagyobb mértékben befolyásolta a vizsgált paramétereket, mint az életkor (Markó–Bóna 2012). A prozódiai szerkezet és a tagolás összefüggéseit is több kutatás elemezte (Gósy 2003; Markó 2010; Váradi 2013). Mády (2012) a fókusz jelölésének prozódiai eszközeit vizsgálta felolvasásban és spontán beszédben. Eredményei szerint a fókusz típusok megkülönböztetésében a következő paraméterek töltenek be elsődleges szerepet: f_0 -minimum és f_0 -maximum, az ezek közötti tartam és az f_0 -maximum pozíciója.

A szintaxis és a prozódia kapcsolatát vizsgálták olvasott és spontán beszédben magyar nyelvre (Szászák–Beke 2012). Az eredmények azt mutatták, hogy a felolvasásban a prozódia alapján a szintaktika magasabb szintjei viselkedhetnek, azonban a rendszer adaptálása spontán beszédre kevésbé volt eredményes (Szászák–Beke 2012). Vizsgálták továbbá, hogy a spontán beszédben prozódiai jellemzőkkel és nem ellenőrzött módszerrel milyen pon-

tossággal lehet automatikusan detektálni az intonációs frázisokat, illetve az azon belüli fonológiai frázisokat. A rendszer nagyon jó minőségben képes ezen frázisokat automatikusan címkézni (Beke et al. 2014).

A beszélő életkora alapvetően meghatározó az alaphangmagasság szempontjából. A gége működése hormonális befolyásoltság alatt áll. A csecsemők gégejének mérete – nemtől függetlenül – nagyjából egyforma, ezáltal a „kisgyermek hangmagassága, hangszíne, hangterjedelme és átlagos beszédhangfekvése lényegében egyforma” (Balázs 1993: 157). Ez a gyermekhang csak a pubertás idejében, hormonális hatásra változik meg; a mutálás eredményeként különül el a férfi és női hang. A nemzetközi kutatások különböző életkorokra teszik ennek idejét; de abban nagyjából megegyeznek az eredmények, hogy a folyamat a lányoknál és a fiúknál nem egy időben zajlik le. 4–6 éves korban még nincs eltérés az f_0 értékében a nemek szerint (Hasek 1980; Nygren et al. 2012); 7–8 éves korban a fiúknál az alaphangmagasság szignifikánsan csökken – a lányoknál már nem változik nagymértékben ebben az életkorban (Ferrand–Bloom 1996). Más kutatások szerint azonban a lányoknál éppen ebben az életkorban megy végbe az alaphangmagasság csökkenése, a fiúknál pedig 8–9 éves kor között zajlik (Hacki–Heitmüller 1999). Egy további vizsgálatban az alaphangmagasság csökkenése a lányoknál 6 és 10 éves kor között, a fiúknál 8 és 10 éves kor között jelentkezett (Whiteside–Hodgson 1999). Mások szerint a mutálás eredményeként bekövetkező alaphangmagasság-változás csak 11 éves korban (Lee et al. 1999), illetve 12 éves korban (Perry et al. 2001) mutatható ki, a női és férfi hang elkülönülése 15 éves korra fejeződik be. Ennek megfelelően a 6–10 éves fiúk és lányok átlagos alaphangmagassága nem mutat szignifikáns különbséget (Sorensen 1989). Guzman és munkatársai (2014) szintén azt találták, hogy a 7–10 éves lányok és fiúk beszédhangjának elkülönítésében az f_0 nem meghatározó tényező. (A nemzetközi szakirodalomban olvasható egymásnak ellentmondó tendenciák természetesen adódhatnak a vizsgálati személyek egyéni jellegzetességeiből.)

Az alaphangszerkezetre vonatkozó kutatások nagyrészt a felnőttek beszédével foglalkoztak. A gyermekek spontán beszédének prozódiai szempontú vizsgálata a közelmúltban indult meg magyar nyelvűre (Markó et al. 2010; Deme 2012; Auszmann–Neuberger 2014; Tóth 2014). Ötéves gyermekeknél még nem volt különbség az f_0 értékében a nemek között (Tóth 2014). A 6–7 éves korosztályban a fiúk átlagos alaphangmagassága szignifikánsan magasabb volt a lányokénál (Deme 2012; Auszmann–Neuberger 2014). A 9 és 11 éves korosztályban nem volt szignifikáns különbség a fiúk és lányok alaphangmagasságának értékében, de a 13 éveseknél már a lányok alaphangja szignifikánsan magasabb volt (Auszmann–Neuberger 2014). Egy másik vizsgálati csoportban a 10 éves fiúk alaphangja szignifikánsan magasabb volt a lányokénál (Tóth 2014). Valószínűsíthető, hogy 10 éves korra a lányok az alaphangmagasság-változás szakaszának végén, míg a fiúk a változás inten-

zívebb periódusában vannak. A változás tehát nem lineáris, és nemenként eltérő ütemben zajlik, ahogy ezt a nemzetközi kutatások is igazolták. A gyermekek fizikai változásai mellett feltételezhető egyéb tényezők – például szocializációs, tanult hatások – befolyása is (Deme 2012; Auszmann–Neuberger 2014; Tóth 2014).

A jelen kutatás célja kisiskolás gyermekek alaphangmagasságának vizsgálata spontán megnyilatkozásokban. A kutatás fő kérdése az volt, hogy az alaphangmagasság értékeiben kimutathatók-e életkori változások az iskolába lépéstől kezdve 3 éven keresztül. Ebben az életkori szakaszban indulhatnak be olyan biológiai folyamatok, amelyek jelentősen befolyásolhatják az alaphangmagasság jellemzőit. A szakirodalom ellentmondó eredményeket tartalmaz arra vonatkozóan, hogy például a nemek tekintetében ezek a paraméterek hogyan változnak a vizsgált életkorban.

Hipotéziseink szerint (i) az alaphangmagasságban tendenciaszerű eltérést lehet adatolni az egyes életkori csoportok között, amely jellemzően a prozódiai eseményekben lesz kimutatható, és nem az alaphangmagasság főbb statisztikai jellemzőiben; (ii) feltételezzük, hogy a vizsgált életkori szakasz végén már statisztikailag is kimutatható különbségeket adatolhatunk a fiúk és a lányok f_0 -értékei között.

Kísérleti személyek, anyag és módszer

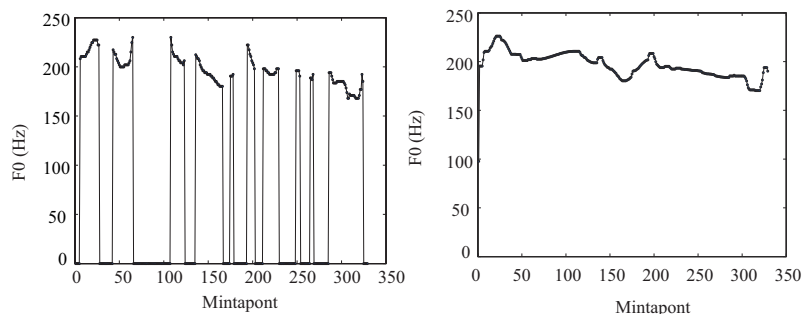
A kutatásban 7, 8 és 9 éves gyermekek vettek részt (korosztályonként 10 fő, 5 lány és 5 fiú minden csoportban). Mindannyian egynyelvűek, ép hallók, nem beszédhibásak, és budapesti iskolába járnak. A gyermekekkel spontán narratívákat rögzítettünk. A protokoll minden esetben ugyanaz volt: a kísérletvezető előre megadta a témákat a gyermekeknek, ezt követően csak akkor szólalt meg, amikor segítő, a közlést továbbvivő kérdésre volt szükség. A narratívák témája a gyermekek családja, lakókörnyezete, hobbjai, iskolai elfoglaltságai voltak. A teljes korpusz időtartama 83 perc (gyermekenként 3–5 perc). A felvételek a megszokott iskolai környezetben, de csendes körülmények között készültek Sony ICD-SX700 típusú hangfelvevővel; a kísérletvezető minden esetben ugyanaz a személy volt. A gyermekek beszédének alaphangszerkezetét összevetettük felnőtt beszélőkével, ehhez a BEA adatbázisból (Gósy et al. 2012) válogattunk narratívákat 10 felnőtt beszélőtől (5 férfi és 5 nő). Az adatközlők életkora 22–35 év. A korpusz időtartama 35 perc (adatközlőnként 3–8 perc).

A korpuszt több szinten annotáltuk a Praat 5.3 programban (Boersma–Weenink 2013). Az annotálást követően kiválogattunk 1015 beszédszakaszt a következő kritériumok mentén: a megnyilatkozás nem tartalmaz zajos részt és irreguláris fonációt, kérdést (a szünettől szünetig tartó megnyilatkozást nevezzük beszédszakasznak).

Az alaphangmagasságot a MATLAB-ban írt YAAP („Yet Another Algorithm for Pitch tracking”) szoftver segítségével nyertük ki (Zahorian–Hu

2008; vö. 1. ábra). A YAAP előnye, hogy robosztus mérési eljárást használ, amely alkalmassá teszi, hogy spontán beszédben mérjünk vele alaphangmagasságot. Mindemellett a tesztelési eredmények is azt mutatták, hogy a YAAP algoritmus szignifikánsan precízebben méri az f_0 -értékeket, mint a Praat (Boersma 1993), RAPT (Talkin 1995) vagy a YIN (Cheveigne-Kawahara 2002).

Az alaphangmagasságot a következő főbb paraméterezéssel végeztük: 25 ms-os Hamming-típusú ablak 10 ms-os tolási értékkel, frekvenciaküszöbök: 60–400 Hz. Mivel az esetünkben folytonos f_0 -görbére volt szükségünk, ezért a YAAP algoritmust úgy használtuk, hogy nem vettük igénybe a zöngétlen részek jelzését. Ebben az esetben az algoritmus kisegítő funkciójával a zöngétlen részek által megszakított görbét inter- és extrapolációval, illetve mediánszűréssel lehet folytonossá tenni. A jelen kutatás során 5 pontos mediánszűrést végeztünk (1. ábra).



1. ábra

Az f_0 reprezentálása utófeldolgozás előtt (balra) és után (jobbra)

Az egyes hangfelvételekben jelölt beszédszakaszok mindegyikében kiszámoltuk a fent bemutatott módon az f_0 -kontúrt, amelyekből statisztikai jellemzőket származtattunk. A statisztikai paraméterek közül a beszédszakasz normál eloszlását közelítő görbe középvértékét: az átlagát; a szóródási mutatóját: az átlagos eltérést; a ferdeséget és a csúcsosságot használtuk. A statisztikai vizsgálatok (varianciaanalízis, t -próba) az SPSS 20.0 programmal történtek. A statisztikai próbákat minden esetben 95%-os konfidenciaintervallum mellett végeztük el.

Prozódiai események detektálása

A prozódiai esemény (PE) minden olyan, az előzményekhez képesti változás, amely egy adaptív küszöbértéket meghalad. Mindezzel a folytonos jelet több kisebb egységre bontjuk úgy, hogy nem egy előre definiált, a spontán beszédre sok esetben rosszul működő szabályrendszert alkalmazunk, hanem a percepció mechanizmushoz közelebb álló eseménydetektálást.

A Kullback–Leibler (*KL*) távolság az egyik leggyakrabban használt algoritmus arra, hogy hogyan mérhető a különbözőség két eloszlás között (Boite–Couvreur 1999). A *KL*-távolságot számos területen alkalmazzák: beszélődetektálás, beszédfelismerés, beszélőfelismerés vagy beszéd vs. nem beszéd detektálás stb. Ezek mellett igen népszerű a *KL*-távolság algoritmus használata szegmentálási problémák megoldására is, mint a beszéd vagy a zene szegmentálása. A jelen tanulmányban a *KL*-távolságot a fonológiai frázisok határainak meghatározására alkalmazzuk. Siegler és munkatársai (1997) kimutatták, hogy a szimmetrikus Kullback–Leibler-távolság egy olyan hatékony távolságmérő eljárás, amely könnyen mérhetővé teszi statisztikailag a különbözőség mértékének kifejezését két beszédjel között. A matematikai háttérét a következőkben ismertetjük: legyen X és Y két random eloszlás, *KL* pedig a különbözőség mértéke e két eloszlás között. A *KL*-ben az eloszlásokat Gauss-eloszlással modellezzük, így az egyes modelleket azok kovarianciamátrixával és a középértékvektorukkal írjuk le.

A Kullback–Leibler-távolság ugyan nemnegatív, de nem valódi metrika, mivel nem szimmetrikus, azaz megkülönböztetheti a modellt és modellezett eloszlást. A *KL* aszimmetrikus távolságot szimmetrikussá lehet tenni a következő lépéssel:

$$KL2(X; Y) = KL(X; Y) + KL(Y; X).$$

Mint korábban írtuk, ha a két eloszlás Gauss-eloszlással közelíthető, akkor a szimmetrikussá tett formában is létezik *KL2* szimmetrikus *KL*-távolság.

A jelen munka során a *KL2*-távolságot a beszédjelben egymást követő részek között számoltuk, amely részek 4 keret hosszúságúnak felelnek meg, vagyis 40 ms időtartamúak (2. ábra). Az ablakhossz 1 keretnyi volt, ami 10 ms-os időtartamnak felel meg. Minden egymást követő beszédész között számolt *KL2*-érték egy folytonos görbét adott.

A következő feladat az volt, hogy *KL2* folytonos jelben megtaláljuk a csúcsokat, amelyek azt jelezték, ahol a két beszédészegmens között a legnagyobb eltérés jelentkezett az akusztikai jellemzők alapján. A magas *KL2*-érték tehát azt feltételezi, hogy a két beszédészegmens között jelentős az eltérés, míg az alacsony *KL2*-érték az azonosságot feltételezi. A csúcsdetektálás szempontjából igen fontos, hogy milyen ablakhosszban keressük az adott csúcsot. Ezért különböző ablakhosszokat alkalmaztunk, amelyeket 10 kerettől (100 ms) 40 keretig (400 ms) növeltünk a *KL2* folytonos görbén.

A csúcsdetektálás szempontjából igen fontos feladat a küszöbérték megválasztása is, mivel ettől függ, hogy az adott *KL2*-értéket váltási pontnak fogadjuk el, vagy sem. Ennek megválasztására két adaptív küszöbölési technikát alkalmaztunk (thr_A és thr_B).

Az első adaptív küszöbértéket úgy számoljuk, hogy az adott keretben található érték középértékét vesszük, majd megszorozzuk egy konstanssal:

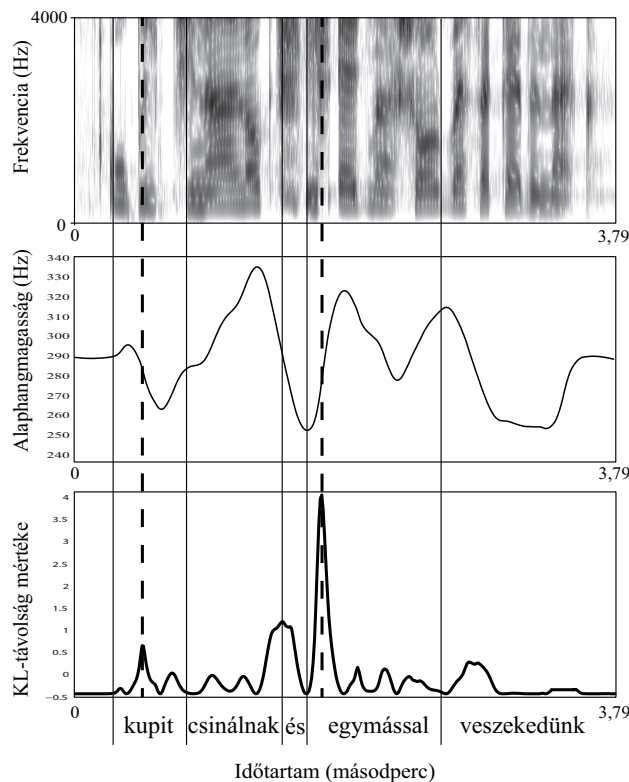
$$thr_{\alpha} = \frac{1}{2N_1} \sum(F),$$

ahol F a jellemzővektor, N_1 az ablak hossza és α a konstans.

A prosódiai események határainak detektálásához az adott értéknek nagyobbak kell lennie, mint thr_B , amelyet a következőképpen számolhatunk:

$$thr_{\beta} = \sigma_F + \beta \frac{1}{2N_1} \sum(F),$$

ahol σ_F az adott ablakhosszban lévő értékek átlagos eltérése, β az ablak hossza.



2. ábra

A prosódiai egységek szegmentálásának folyamata

Az első küszöbérték azt biztosítja, hogy az adott érték nagyobb, mint a környező területen számított értékek, amely egy rövid idejű ablakra számolandó. A második küszöbérték – amelyet egy hosszabb ablakra számolunk – kezeli

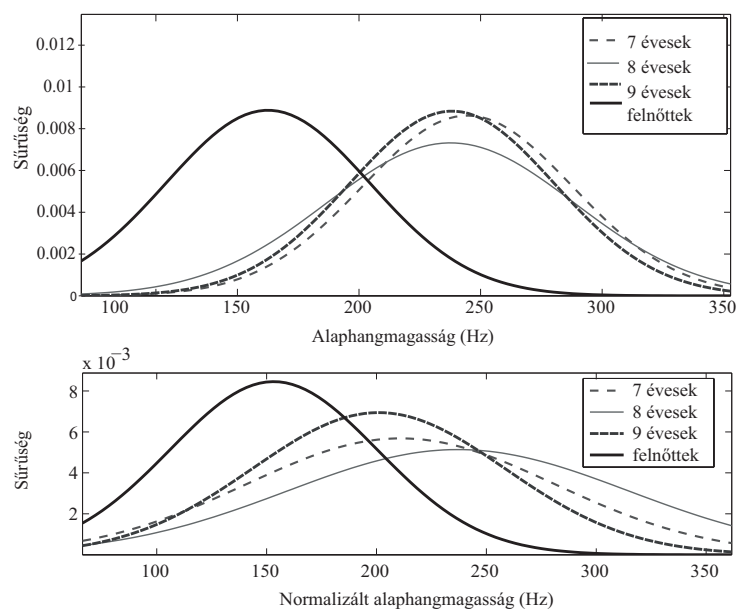
az általános tendenciákat az ablakon kívüli adatok változásának figyelembe vételével.

Az ablakok méretét 3 és 4 másodpercre állítottuk, ennek a küszöbölési technikának a használata biztosította, hogy a téves elfogadások száma csökkenjen, és csak a valóban magas $KL2$ -értékek legyen elfogadva, amelyek a prozódiai események határait jelentették.

Eredmények

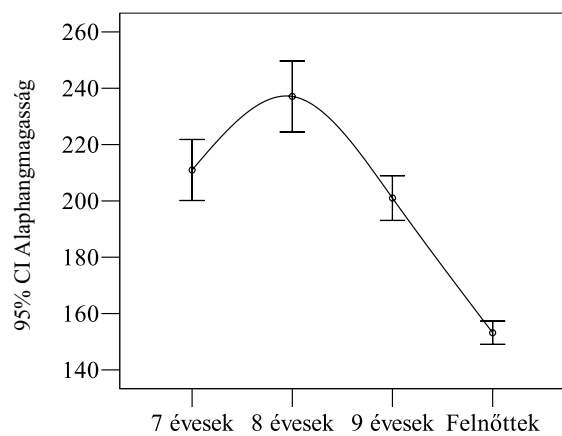
A több szempontos varianciaanalízis szerint a beszélő neme meghatározza az alaphangmagasság átlagos értékét [$F(1, 10) = 6,446$; $p = 0,016$; $\eta^2 = 0,168$]; a 95%-os konfidenciaintervallum alsó értékét [$F(1, 10) = 6,614$; $p = 0,015$; $\eta^2 = 0,171$] és felső értékét [$F(1, 10) = 6,282$; $p = 0,017$; $\eta^2 = 0,164$]. A csoport-szintű elemzéshez t -próbát használtunk. Az eredmények szerint a 7 éveseknél nincs szignifikáns különbség az átlagos alaphangmagasság értékében. A fiúk f_0 -átlaga 241 Hz (átl. elt.: 19 Hz), a lányoké valamivel alacsonyabb, 236 Hz (átl. elt.: 21 Hz). A 8 éves korosztályban sincs szignifikáns különbség a vizsgált csoportban a fiúk és lányok alaphangmagassága között. A fiúk f_0 -átlaga 232 Hz (átl. elt.: 22 Hz), a lányoké 235 Hz (átl. elt.: 24 Hz). A 9 éveseknél sem különbözik egymástól szignifikáns mértékben a lányok és a fiúk alaphangmagassága, noha kicsit nagyobb eltérés volt adatolható a két csoport között, mint a fiatalabb korosztályokban. A fiúk f_0 -átlaga 224 Hz (átl. elt.: 8 Hz), a lányoké 239 Hz (átl. elt.: 16 Hz). Az elvártaknak megfelelően a felnőtt férfiak és nők alaphangmagassága szignifikánsan különbözik, az átlagok közötti eltérés 40 Hz ($t = -8,779$; $p < 0,001$). A férfiaknál mért átlagérték 133 Hz (átl. elt.: 6 Hz), a nőknél pedig 176 Hz (átl. elt.: 8 Hz). A kutatás fő szempontja a beszélő életkorának hatása volt az f_0 -ra, ezért az adatokat úgy normalizáltuk a későbbi elemzésekre, hogy eltűnjenek a beszélők neméből adódó, illetve az egyéni ejtésből fakadó különbségek. Az adatokat csoportszinten normalizáltuk két lépésben. A z -normalizálás után az akusztikai jellemzőt újra skáláztuk a teljes populációban mért akusztikai jellemző minimum és maximum értékére. Az így normalizált adatok voltak a korcsoportokra vonatkozó statisztikai elemzés bemenetei, az eljárás eredményét a 3. ábra mutatja.

Az egyes beszélőkre meghatározott átlagos alaphangmagasság normalizált értéke a 7 éveseknél 210 Hz (átl. elt.: 70 Hz), a 8 éveseknél 237 Hz (átl. elt.: 77 Hz), a 9 éveseknél 201 Hz (átl. elt.: 57 Hz). A felnőttek csoportjában 153 Hz volt az f_0 átlagos értéke (átl. elt.: 47 Hz). A több szempontos varianciaanalízis szerint a kor hatással van a beszélő alaphangmagasságára [$F(3, 1015) = 104,502$; $p < 0,001$; $\eta^2 = 0,236$]. Szignifikáns különbség volt igazolható a felnőttek és mindhárom kisiskolás csoport között (minden esetben $p < 0,001$, vö. 4. ábra). A 8 évesek átlagos alaphangmagassága szignifikánsan magasabb, mint a 7 éveseké és a 9 éveseké (mindkét esetben: $p < 0,001$).



3. ábra

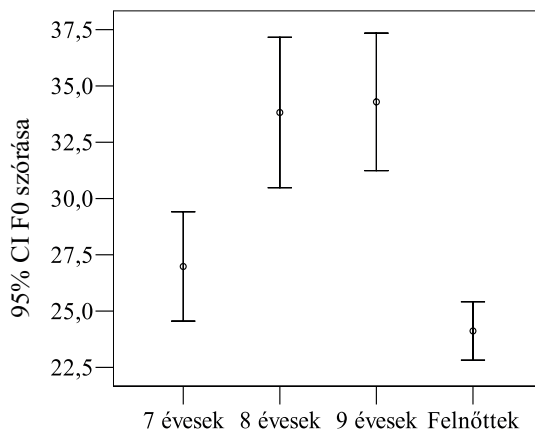
Az alaphangmagasság (fent) és a normalizált alaphangmagasság (lent) értékei az egyes korcsoportokban



4. ábra

Az f_0 középértéke az életkor függvényében

Az f_0 szórása szignifikáns eltérést mutat a beszélő életkorának függvényében [$F(3, 1015) = 22,602$; $p < 0,001$; $\eta^2 = 0,063$]. A szórás értéke a 7 éveseknél 26 Hz (átl. elt: 15 Hz), a 8 éveseknél 33 Hz (átl. elt: 20 Hz), a 9 éveseknél 34 Hz (átl. elt: 22 Hz), a felnőtteknél pedig 24 Hz (átl. elt: 20 Hz). A 8 és 9 éveseknél meghatározott szórásértékek szignifikánsan nagyobbak a felnőtteknél (mindkét esetben $p < 0,001$). A szórás értékében a 8 és 9 évesek kivételével szignifikáns a különbség a kisiskolás csoportok között (minden esetben $p < 0,001$). Az f_0 variabilitása tehát változik az életkorral, és felnőttkorban jóval kevésbé szórnak az értékek, mint a gyermekeknél (5. ábra).



5. ábra

Az f_0 szórása az egyes korosztályokban

Az f_0 -eloszlásra illesztett normál eloszlás ferdesége a 7 éveseknél 0,1 (átl. elt: 1,1), a 8 éveseknél $-0,4$ (átl. elt: 1,3), a 9 éveseknél $-0,03$ (átl. elt: 2,06), a felnőtteknél pedig $-0,1$ (átl. elt: 1,5). Az egyes életkori csoportoknál meghatározott f_0 eloszlása balra ferde, amely arra utal, hogy az átlagos alaphangmagasság-értékhez képes az alsóbb frekvenciaértékek súlya nagyobb. Az egyes korosztályok között nem volt szignifikáns eltérés.

Az f_0 -eloszlásra illesztett normál eloszlás csúcsossága a 7 éveseknél 3,6 (átl. elt: 2,03), a 8 éveseknél 5,2 (átl. elt: 3,4), a 9 éveseknél 8,4 (átl. elt: 7,3), a felnőtteknél 5,7 (átl. elt: 4,2). A különbség statisztikailag szignifikáns az egyes korosztályok között [$F(3, 1015) = 33,865$; $p < 0,001$; $\eta^2 = 0,091$]. Mindhárom kisiskolás csoport értéke szignifikánsan különbözik a felnőtteké-től (minden esetben $p < 0,001$). A 8 évesek ebben a jellemzőben is különböznek a 7 és 9 évesektől (mindkét esetben: $p < 0,001$).

Elemeztük a hangköz értékeit is, vagyis az f_0 maximumának és minimumának hányadosát. A hangköz értéke 7 éveseknél 1,43 (átl. elt: 0,33), a 8 éveseknél 1,5 (átl. elt: 0,37), a 9 éveseknél 1,48 (átl. elt: 0,4), a felnőtteknél 1,44 (átl. elt: 0,22). Ebben a paraméterben tehát nincs különbség a beszélő életkorának függvényében.

Elemeztük a megnyilatkozások f_0 -kontúrját, ezt követően k -közép algoritmussal (a minták közötti hasonlóságot korrelációval mérve) a kontúrokat csoportokra bontottuk. Ezzel az eljárással megkaptuk a tipikus dallammentekeket. A klaszterek számát (jelen esetben a tipikus kontúrok) automatikusan választottuk meg a silhouette analízissel (Rousseeuw 1987). A silhouette eljárással a k -közép csoportba való sorolásának erősségét lehet vizsgálni, illetve hogy az egyes elemek a csoportban milyen erős tagsággal rendelkeznek. A silhouette analízis levezetése a következőképpen történik:

$$SW_i = (b_i - a_i) / \max(a_i - b_i),$$

ahol a az átlagos távolság az i -edik pont és az i -edik pont klaszterének többi egyede között, míg b az átlagos távolság az i -edik pont és a többi klaszter egyedei között, így minden esetben: $-1 < SW_i < 1$. Az SW_i az alábbiak szerint értelmezhető (1. táblázat):

1. táblázat: A silhouette analízis értékei és a hozzájuk tartozó kategóriák

A silhouette elemzés kiértékelése	
SW_i értéke	Kapcsolat erőssége
0,71–1,00	erős
0,51–0,70	mérsékelt
0,26–0,50	gyenge
$\leq 0,25$	nincs

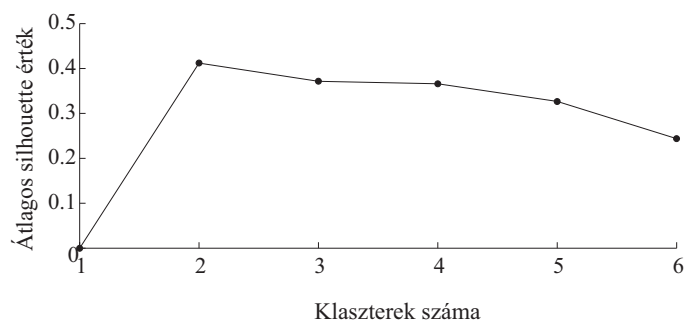
A csoportok számának megállapítására a silhouette eljárással kapott SW_i -értékek középértékét vettük, így jellemezve a klaszterezés eredményességét (6. ábra). Az ábrán látható, hogy a silhouette átlagos értéke akkor volt a legmagasabb, ha a csoportok számát kettőre választottuk.

A tipikus menetek két csoportba voltak sorolhatóak. Az első típus esetén főként emelkedő tendenciát mutat az f_0 , míg a kettes kontúr típus inkább eső tendenciát mutat. A gyermekek esetében ez a két kontúrtípus jól elkülöníthető, erősebb eséssel vagy emelkedéssel valósul meg, amely főként a kontúr vége felé adatható. A felnőttek kontúrja a gyermekekéhez képest inkább lebegő típusú, és a különbség főként a kontúr elején látható (vö. 7. ábra).

A kontúrtípusok előfordulási arányát a korpuszban a 2. táblázat mutatja.

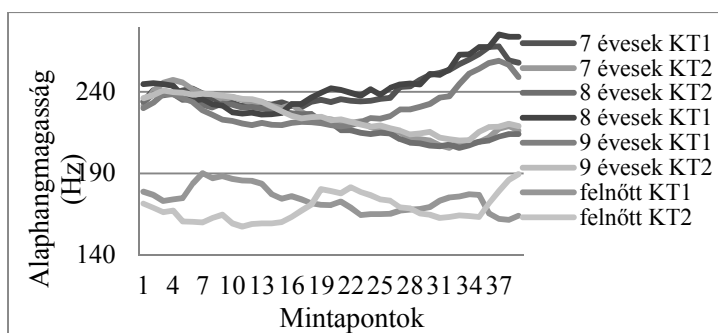
Meghatároztuk az egyes típusok meredekségét; az adatokra statisztikai vizsgálatokat végeztünk. Az 1. kontúrtípus meredekségét meghatározza a beszélő életkora [$F(3, 281) = 7,766$; $p < 0,001$; $\eta^2 = 0,077$], statisztikailag szignifikáns különbséget ugyanakkor csak a felnőttek és a kisiskolás csoportok

között adatoltunk (minden esetben $p < 0,001$). Az első kontúrtípus meredeksége a 7 éveseknél 0,71 (átl. elt: 0,83), a 8 éveseknél 0,76 (átl. elt: 1,08), a 9 éveseknél 0,83 (átl. elt: 1,17), a felnőtteknél pedig $-0,4$ (átl. elt: 0,55).



6. ábra

A silhouette érték a klaszterek számának függvényében



7. ábra

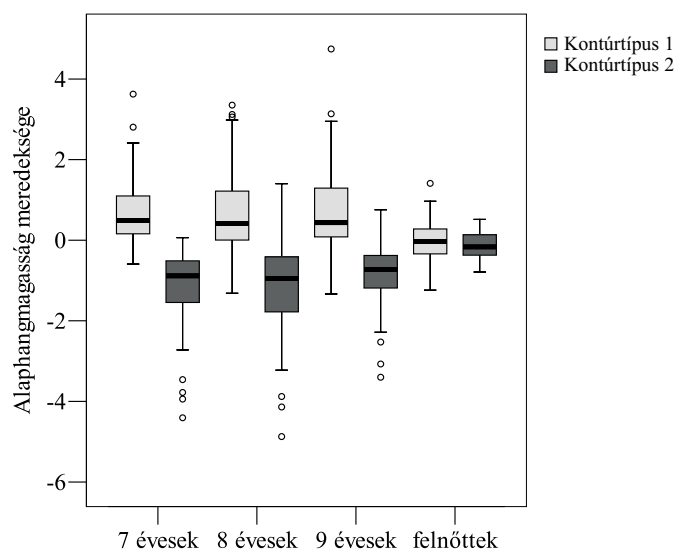
Az f_0 fő kontúrjai az egyes korosztályokban (KT = kontúrtípus)

2. táblázat: Az f_0 -kontúrtípusok előfordulása az egyes korcsoportokban

Korosztály	Az f_0 -kontúrtípusok előfordulása (%)	
	1. típus	2. típus
7 évesek	59	41
8 évesek	74	26
9 évesek	68	32
Felnőttek	64	36

A 2. kontúrtípus meredekségét szintén meghatározza a beszélő életkora [$F(3, 315) = 7,701$; $p < 0,001$; $\eta^2 = 0,083$], statisztikailag szignifikáns különbséget ennél a típusnál is csak a felnőttek és a kisiskolás csoportok között adatoltunk (minden esetben $p < 0,001$). A második kontúrtípus meredeksége a 7 éveseknél $-1,10$ (átl. elt: $0,88$), a 8 éveseknél $-1,17$ (átl. elt: $1,05$), a 9 éveseknél $-0,93$ (átl. elt: $0,89$), a felnőtteknél pedig $-0,14$ (átl. elt: $0,33$).

Elemeztük azt is, hogy az egyes korosztályokban az 1. vagy a 2. kontúrtípus valósul-e meg nagyobb meredekséggel (8. ábra). A kontúr típusa meghatározza az f_0 -változás meredekségét [$F(1, 596) = 255,692$; $p < 0,001$; $\eta^2 = 0,303$]. A gyermekek esetében minden korcsoportban szignifikáns a különbség a két kontúrtípus között ($p < 0,05$), a felnőtteknél ez az eltérés azonban matematikailag nem igazolható.



8. ábra

Az f_0 fő kontúrjai az egyes korosztályokban

Következtetések

A kutatás fő kérdése az volt, hogy miként változik az alaphangmagasság paramétereinek értéke a beszélő életkorának függvényében. A felnőttek átlagos alaphangmagassága természetesen szignifikánsan alacsonyabb volt a gyermekekéhez képest; de a kisiskolás csoportok f_0 -értékeiben is találtunk különbséget. A 8 évesek átlagos alaphangmagassága szignifikánsan magasabb volt a jelen vizsgálati csoportban, mint a náluk egy évvel fiatalabb, illetve egy évvel idősebb gyermekeké. Az f_0 -értékek szórása a felnőtteknél szig-

nifikánsan kisebb, mint a gyermekeknél, mivel a gyermek beszédében a zöngémodulálása kevésbé stabil a felnőttekéhez képest.

Mindezek mellett vizsgáltuk a nem hatását az alaphangmagasságra. Az eredmények azt mutatták, hogy kisiskolás korban a jelen vizsgálati csoportban még nem lehetett szignifikáns különbséget kimutatni a lányok és a fiúk alaphangmagassága között. A tendencia azt mutatta, hogy a 7 éveseknél a lányok átlagos alaphangmagassága kicsit alacsonyabb a fiúkénál, a 8 éveseknél gyakorlatilag nincs különbség a két csoport között; a 9 éveseknél pedig már a lányok átlagos f_0 -értéke magasabb a fiúkénál.

Az f_0 -kontúrok esetében minden korosztályban az emelkedő jellegű gyakoribb, mint az eső jellegű. Mindkét típusnak a meredeksége szignifikánsan nagyobb a gyermekeknél, mint a felnőttek beszédében, tehát az alaphangmagasság nagyobb változásokat mutat egy megnyilatkozáson belül a gyermekek esetében. A gyermekeknél jellemzőbb az emelkedés a kontúr vége felé. Ez a jellegzetes dallamemelés hozzájárul a gyermekek beszédéről kialakult perpciós benyomáshoz.

Irodalom

- Ausmann, Anita – Neuberger, Tilda 2014. Age- and gender-related differences in formant structure during the stabilization process of vowels. In: *Proceedings of the Olomouc Linguistics Colloquium 2014*. 663–676.
- Balázs Boglárka 1993. Az időskori hangképzés jellemzői. *Beszédkutatás* 1993. 156–165.
- Beke András 2008a. A felolvasás és a spontán beszéd alaphangszerkezeteinek vizsgálata. *Beszédkutatás* 2008. 93–107.
- Beke András 2008b. Az alaphangfrekvencia-eloszlás modellezése a beszélőfelismeréshez. *Alkalmazott Nyelvtudomány* 8/1–2. 121–133.
- Beke, András – Szaszák, György – Váradi, Viola 2013. Automatic phrase segmentation and clustering in spontaneous speech. In: *Proceedings of IEEE 4th International Conference on Cognitive Infocommunications, CogInfoCom 2013*. Budapest. 452–462.
- Boersma, Paul 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In: *Proceedings of the Institute of Phonetic Sciences 17*. University of Amsterdam. 97–110.
- Boersma, Paul – Weenink, David 2013. Praat: doing phonetics by computer. <http://www.fon.hum.uva.nl/praat/> (A letöltés ideje: 2013. október 10.)
- Boite, Jean-Marc – Couvreur, Laurent 1999. Speaker tracking in broadcast audio material in the frame work of the THISL project. In: *Proceedings of the workshop on accessing information in spoken audio (ESCAETRW99)*. 84–89.
- Bolla Kálmán 1992. 1992. *Szupraszegmentális elemzések*. Egyetemi Fonetikai Füzetek 7. ELTE Fonetikai Tanszék, Budapest.
- de Cheveigne, Alain – Kawahara, Hideki. 2002. YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustic Society of America* 111. 1917–1930.

- Deme Andrea 2012. Óvodások magánhangzóinak akusztikai jellemzői. In Markó Alexandra (szerk.): *Beszédtudomány: az anyanyelv-elsajátítástól a zöngékezdési időig*. ELTE–MTA, Budapest. 77–99.
- Ferrand, Carole T. – Bloom, Ronald R. 1996. Gender differences in children's intonational patterns. *Journal of Voice* 10/3. 284–291.
- Gósy Mária 2003. Virtuális mondatok a spontán beszédben. *Beszédkutatás* 2003. 19–43.
- Gósy, Mária – Terken, Jacques 1994. Question marking in Hungarian: timing and height of pitch peaks. *Journal of Phonetics* 22. 269–281.
- Gósy Mária – Gyarmathy Dorottya – Horváth Viktória – Grácz Tekla Etelka – Beke András – Neuberger Tilda – Nikléczy Péter 2012. BEA: Beszélt nyelvi adatbázis. In Gósy Mária (szerk.): *Beszéd, adatbázis, kutatások*. Akadémiai Kiadó, Budapest. 9–24.
- Guzman, Marco 2014. Acoustic markers to differentiate gender in prepubescent children's speaking and singing voice. *International Journal of Pediatric Otorhinolaryngology* 78. 1592–1598.
- Hacki, Tamas – Heitmüller, S. 1999. Development of the child's voice: premutation, mutation. *International Journal of Pediatric Otorhinolaryngology* 49. 141–144.
- Hasek, Carol S. – Singh, Sadanand – Murry, T. 1980. Acoustic attributes of preadolescent voices. *Journal of the Acoustical Society of America* 68. 1262–1265.
- Lee, Sungbok – Potamianos, Alexandros – Narayanan, Shrikanth 1999. Acoustics of children's speech: developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America* 105/3. 1455–1468.
- Mády Katalin 2012. A fókusz prosódiai jelölése felolvasásban és spontán beszédben. In Gósy Mária (szerk.): *Beszéd, adatbázis, kutatások*. Akadémiai Kiadó, Budapest. 91–107.
- Markó Alexandra 2005. *A spontán beszéd néhány szuprasegmentális jellegzetessége. Monologikus és dialogikus szövegek összevetése, valamint a hűmmögés vizsgálata*. PhD-disszertáció. ELTE, Budapest.
- Markó Alexandra 2009. Stigmatizált hanglejtésforma a spontán beszédben. *Beszédkutatás* 2009. 88–106.
- Markó Alexandra 2010. A prosódia szerepe a spontán beszéd tagolásában. *Beszédkutatás* 2010. 82–99.
- Markó Alexandra – Grácz Tekla Etelka – Imre Angéla 2010. A diskurzusjelölők használatának fejlődése: a hűmmögés formai és funkcionális sajátosságai különböző életkorokban. In Navracsics Judit (szerk.): *Nyelv, beszéd, írás. Pszicholingvisztikai tanulmányok*. Tinta Kiadó, Budapest. 82–92.
- Markó Alexandra – Bóna Judit 2012. Eltérő beszédmódok intonációs sajátosságai fiatal és idős korban. In Balázs Géza – Veszelszki Ágnes (szerk.): *Nyelv és kultúra – kulturális nyelvészet*. Magyar Szemiotikai Társaság, Budapest. 253–258.
- Nygren, Mariana – Tyboni, Mikaela – Lindström, Frederic – McAllister, Anita – van Doorn, Jan 2012. Gender differences in children's voice use in a day care environment. *Journal of Voice* 26/6. 817–e15–18.
- Olaszy Gábor 1995. A kérés, a figyelmeztetés, a felszólítás és a kérdés prosódiaja a kijelentő mondat tükrében. *Beszédkutatás* 1995. 46–61.
- Olaszy Gábor 2002. The most important prosody patterns of Hungarian. *Acta Linguistica Hungarica* 49/3–4. 277–306.

- Perry, Theodore L. – Ohde, Ralph N. – Ashmead, Danieal H. 2001. The acoustic bases for gender identification from children's voices. *Journal of the Acoustical Society of America* 109/6. 2988–2998.
- Rousseeuw Peter J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Computational and Applied Mathematics* 20. 53–65.
- Siegler, Matthew A. – Jain, Uday – Raj, Bhiksha – Stern, Richard M. 1997. Automatic segmentation, classification, and clustering of broadcast news audio. In: *Proceeding of the DARPA speech recognition workshop*. Chantilly, Virginia. 97–99.
- Sorensen, David N. 1989. A fundamental frequency investigation of children ages 6–10 years old. *Journal of Communication Disorders* 22/2. 115–123.
- Szaszák György 2008. *A szuprasegmentális jellemzők szerepe és felhasználása a beszédfelismerésben*. PhD-disszertáció. BME, Budapest.
- Szaszák, György – Beke, András 2012. Exploiting prosody for syntactic analysis in automatic speech understanding. *Journal of Language Modelling* 0/1.143–172.
- Talkin, David 1995. A robust algorithm for pitch tracking RAPT. In Kleijn, W. Bastiaan – Paliwal, Kuldip K. (eds.): *Speech coding and synthesis*. Elsevier Science, New York. 495–518.
- Tóth Andrea 2014. Gyermekek nemének és életkorának meghatározása a beszédük alapján. *Beszéd kutatás 2014*. 98–111.
- Váradi Viola 2013. *A spontán beszéd szegmentálása produkciós és percepciós szempontból*. PhD-disszertáció. ELTE, Budapest.
- Varga László 1988. A gazdagréti kábeltelevízió műsorából válogatott anyag intonációs átirata. In Kontra Miklós (szerk.): *Beszélt nyelvi tanulmányok*. Linguistica A/1. MTA Nyelvtudományi Intézet, Budapest. 5–21.
- Varga László 1994. A hanglejtés. In Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 2. Fonológia*. Akadémiai Kiadó, Budapest. 468–549.
- Varga László 2002. *Intonation and stress. Evidence from Hungarian*. Palgrave Macmillan, Houndmills, Basingstoke.
- Whiteside, Sandra P. – Hodgson, Carolyn 1999. Acoustic characteristics in 6–10-year-old children's voices: some preliminary findings. *Logopedics Phoniatrics Vocology* 24. 6–13.
- Zahorian, Stephen A. – Hu, Honbing 2008. A spectral/temporal method for robust fundamental frequency tracking. *Journal of the Acoustic Society of America* 123/6. 4559–4571.

A tanulmány a Bolyai János Kutatási Ösztöndíj és az OTKA 108762 számú pályázat támogatásával készült.

SZÓ ELEJI ZÁRHANGOK ZÖNGEKEZDÉSI IDEJE: BESZÉDPRODUKCIÓS ADATOK AZ ÓVODÁS KOROSZTÁLYRA VONATKOZÓLAG

Tar Éva

Bevezetés

A zöngékezdési idő (az angol voice onset time rövidítéséből: VOT) a zárhangok artikulációjához kapcsolódó idői paraméter; a szájüregi zár feloldásának kezdete és a hangszalagrezgés megindulása közti időintervallumot jelöli. Az artikulációs és fonációs működés a toldalékcsoben áramló levegő tulajdonságait befolyásolja, a nyomásváltozás az akusztikai jelben megjelenik, a regisztrátumon mérhetővé válik. A zöngé megindulásának a zárfelpattanáshoz való időbeli viszonya alapján a VOT különböző típusait különböztetik meg. Amennyiben a hangszalagok rezgése a zárfelpattanás előtt indul, ún. előzöngé valósul meg, a VOT értéke negatív lesz. A zárfelpattanás utáni zöngéindulás esetében a zárhangokat a késés időtartamától függően legáltalánosabban további két fonetikai kategóriába sorolják: aspirálatlan zárhang (rövid pozitív VOT) és aspirált zárhang (hosszú pozitív VOT) (Lisker–Abramson 1964).

A magyar nyelvben az obstruensek közös tulajdonsága, hogy elkülöníthetők a zöngésségi oppozíció mentén (Siptár–Törkenczy 2007). A zöngés explozívak fonetikailag előzöngével, a zöngétlenek rövid pozitív VOT-val valósulnak meg (Gósy 2004). A VOT értéke szó eleji pozícióban, felnőttek szölista-felolvasása révén nyert adatai alapján a következőképp alakult: $p = 9,7$ ms, $t = 16$ ms, $k = 37,6$ ms; $b = -94,6$, $d = -95,1$ ms, $g = -89,6$ ms (Gósy–Ringen 2009). A zöngétlen hangok VOT-értékeit a képzés helye befolyásolta, a hátrébb képzett az előrébb képzettnél szignifikánsan hosszabb zöngékezdési idővel realizálódott; valamint a zöngés célfonémák mindegyike előzöngével valósult meg. Felnőtt beszélőkkel folytatott kutatások szerint a zöngékezdési idő tartamára további nyelvi és nyelven kívüli tényezők is hatnak: a zárhangot követő magánhangzó minősége (Gósy 2000), az életkor (Bóna 2011), a beszéd típus (Gósy 2001) és az egyéni sajátosságok (Neuberger 2014). Továbbá Neuberger és Grácsi (2013) nemek közti eltérést igazolt a felpattanás időtartama és az explozíva teljes időtartama arányát illetően; Grácsi és Kohári (2002) pedig azt találta, hogy a többszörös felpattanások gyakoriságát (a képzési helyen kívül) az egyéni stratégia is befolyásolja.

A fonológiai fejlődés folyamán a gyermek elsajátítja a zöngésségi kontrasztra vonatkozó nyelvi tudást, valamint azokat a készségeket, melyek a megvalósítás alapjában állnak (az artikulációs, laringális és légzési szakasz-

ban zajló működések nyelvspecifikus koordinálását). A fejlődés éveikig tartó folyamat. Észlelésalapú vizsgálatok nyomán tudjuk, hogy magyar anyanyelvű gyermekek beszédprodukciónak a fejlődés korai időszakában megjelenik a zöngésségi oppozíció (S. Meggyes 1971; Gósy 1984, 1998). A hibaelemzések azonban azt mutatják, hogy szóhatár pozícióban gyakran előfordul a zöngés obstruensek (köztük is leginkább a zárhangok) zöngétlenítése (Sebestyén 2006). Továbbá, ez a fonetikai pozíciótól függő hibamintázat tartósan fennmarad, és gyakori jelenség fonológiai zavar esetén is (Sebestyén 2008). A zöngésség tekintetében a nemek fejlődést befolyásoló hatására a magyar nyelvre vonatkozóan nincs adat. Egyéb nyelvek vizsgálatok a nemek közti különbség tekintetében ellentmondóak a vizsgálati eredmények. Néhány szerző nemek közti különbségről tudósít (Smit et al. 1990; So 2006), mások (Dodd et al. 2003) nem találtak eltérést a két nem közt a zöngésség elsajátítása tekintetében.

Akusztikai (leginkább VOT-re vonatkozó) elemzések tovább árnyalták a szó eleji explozívák zöngéssége fejlődésére vonatkozó tudást. A fonológiai oppozíció megvalósulási alapjában álló fonációtípusok eltérőek a létrehozásához szükséges motoros kontroll tekintetében, az elsajátítás mintázata a leginkább igénybevevő (és így legkésőbb kialakuló) felé haladva a következő: rövid pozitív VOT > hosszú pozitív VOT > előzöng (negatív VOT) (pl. Gandour et al. 1986). Az előzöngével képzett zárhangok a legkomplexxebbek a motoros koordináció szempontjából (Kewley-Port–Preston 1974). A gyermeknek egyrészt meg kell tanulnia a két beszédprodukciónak a rendszer (a gége és a szájüregi artikulációs szervek) működését nyelvspecifikus módon időben összerendezni, másrészt el kell sajátítania további olyan artikulációs stratégiá(ka)t, melyek biztosítják a fonációhoz szükséges aerodinamikai feltételeket az előzöngé produkciójakor. A zöngé fenntartása a zárszakasz alatt ugyanis, legalábbis a felnőtteknél mért időtartamban, az artikulációs szervek aktív működésével érhető el. Különböző stratégiával lehet a zöngé fenntartásához szükséges aerodinamikai feltételeket javítani, a szájüregi nyomást csökkenteni (pl. Ohala 2011). Magyar anyanyelvű beszélőknél szó eleji helyzetben Grácsi (2012) svával történő indítást figyelt meg. Atipikus nyelvi fejlődésű gyermekek több esetben nazálisbetoldást vagy nazalizálást alkalmaztak (S. Tar 2013).

A fokozatos fejlődést egy más szempontból közelítve Scobbie és munkatársai (2000) a felnőttszerű kontraszt elsajátításának négy fejlődési szakaszát különböztetik meg: hiányzó kontraszt, felnőtt által nem észlelhető fonetikai különbség, éretlen kontraszt, érett kontraszt. A felnőttszerű VOT-érték az észlelhető különbség megjelenése után évekkal lesz csak jellemző a gyermek produkciójára.

Az életkoron kívül a nemek hatását is több kutatás vizsgálta. Nemek közti különbséget a VOT-elemzések leginkább serdülőkor után dokumentálnak, amikor az eltérés biológiai okokra (pl. gége dimorfizmusa) vezethető vissza (vö. Koenig 2000). A gyermekkori nemek közti különbségre irányuló kutató-

sok száma viszonylag csekély. Whiteside és Marshall (1998), Whiteside és munkatársai (2004) és Karlsson és munkatársai (2004) a nem és az életkor interakcióját találták a VOT időtartamára, valamint azt, hogy ez a hatás csak bizonyos fonémáknál és bizonyos fonetikai környezetben érvényesül. A szerzők biológiai tényezőknek és egyéni fejlődési útból fakadó eltéréseknek tulajdonították az eredményeket. Whiteside és Marshall (1998), valamint Nissen és Fox (2009) viszont azt veti fel, hogy a nemek közti különbség bizonyos mintázata (pl. a zöngétlen/zöngés párok közti jelentősebb elkülönülés a VOT-dimenzióban) szociofonetikai hatások következménye is lehet.

Magyar nyelvű gyermekekre vonatkozóan Bóna és Auszmann (2014) végzett vizsgálatokat. A szerzők 9, 11 és 13 éves gyermekek spontánbeszéd-mintáiban elemezték a *p*, *t*, *k* zöngekezdési időtartamát. A három fonetikai helyzetben nyert adatokat együtt elemezve azt találták, hogy az életkor szignifikánsan befolyásolta a VOT-k értékét, fonémánként azonban eltérő irányban: *p* és *k* esetén az életkorral csökkent, a *t* esetében nőtt az időtartam. A képzés helye alapján a VOT-értékek szignifikánsan különböztek, kivéve a *p* és *t* közötti különbséget 9 éves korban. A nemenkénti elemzés életkoronként és fonémánként eltérő eredményt adott: a *p*-re nem mutatkozott különbség, a *t*-re a két idősebb korcsoportban, míg a *k*-ra a legfiatalabb korosztályban találtak nemenkénti eltérést. Felnőttszerű értéket a VOT-k különböző életkorban értek el: $k > p > t$.

A jelen tanulmány egy szélesebb, a 3–6 éves korosztály produktív fonológiai tudásának feltárására irányuló kutatás részeként mutatja be a zöngésségre vonatkozó adatokat. Célja a szó eleji explozívák zöngekezdési idejében a 3;0–3;5, 3;6–3;11 és 5;6–5;11 életkorban azonosítható sajátosságok bemutatása. A két fiatalabb csoport különválasztását azok a kutatási eredmények motiválták, melyek a zöngésségi kontraszt elsajátításában a 3–4 éves életkori szakaszban találtak nemenkénti különbséget (pl. Smit et al. 1990). Az idősebb korcsoport kiemelt életkori kategóriaként kezelését az a tanulmány alapozza meg, amely szerint 5;6–5;11 éves korra a magyar anyanyelvű gyermekek elsajátítják a fonológia szegmentális rendszerét (Sebestyén 2006). A jelen tanulmány alapján álló kutatásban arra kerestük a választ, hogy hogyan alakul különböző nyelvi és nem nyelvi tényezők hatására (i) az előzöngés explozívák gyakorisága, illetőleg (ii) a VOT időtartama. Figyelembe véve a fiatalabb és idősebb gyermekek közti nyelvfelépésbeli különbséget, a fejlődési aspektust az elől képzett zárhangokra, illetőleg a pozitív VOT-k időtartamára vonatkozólag vizsgáltuk. Hogy a vizsgált korosztályról teljesebb képet nyerjünk, az 5;6–5;11 éves gyermekek beszédmintáján a velárisokra és a negatív VOT időtartamára vonatkozóan további elemzéseket végeztünk. Előfeltevéseket a nemek hatása tekintetében fogalmaztunk meg. Azt vártuk, hogy (i) az előzöngé gyakorisága a 5;6–5;11 éves korcsoportban nem különbözik a fiúk és a lányok mintájában, továbbá (ii) a VOT időtartamában nemek közti eltérés – ha fennáll – csak bizonyos fonémák és bizonyos életkori csoport esetében igazolható.

Anyag, módszer, vizsgálati személyek

A beszédminta kiváltása képmegnevezéssel történt. A jelen kutatásban elemzett adatokat egy, a fonológiai profil feltárásához összeállított, egymorfémás főneveket tartalmazó szólista (vö. Sebestyenné 2006) 22 tétele adta (1. táblázat). A zöngétlen expozíciók kiváltásához öt, a zöngésekéhez legalább három szót használtunk. Minden sorozat tartalmazott egy olyan szóalakat, melyben a szó eleji zárhangot felső állású magánhangzó követte; a célhang fonetikai környezetét egyéb tekintetben nem kontrolláltuk.

1. táblázat: Szólista

p	b	t	d	k	g
<i>pízsama</i>	<i>busz</i>	<i>tű</i>	<i>dinnye</i>	<i>kutya</i>	<i>gép</i>
<i>pötty</i>	<i>bodza</i>	<i>tehén</i>	<i>doboz</i>	<i>kés</i>	<i>gomb</i>
<i>pohár</i>	<i>bohóc</i>	<i>templom</i>	<i>dominó</i>	<i>könyv</i>	<i>golya</i>
<i>perec</i>	<i>bélyeg</i>	<i>telefon</i>	<i>darázs</i>	<i>kard</i>	
<i>papucs</i>	<i>bagoly</i>	<i>táska</i>		<i>karácsony</i>	

A digitálisan (Sony ICD–MS525 típusú diktafonnal) rögzített hanganyagot konvertálás után a Praat 5.3 szoftverrel (Boersma–Weenink 2011) elemeztük. Az adatolást manuálisan, a rezgéskep és a spektrogram alapján végeztük. A hangszínképelemzéshez széles sávú spektrogramot használtunk, 5000 Hz-es tartományban.

Az **előzöngé gyakorisága** esetében az **életkor**, a **nem**, a **képzés helye**, valamint a **fonológiai zöngesség** hatását vizsgáltuk. Az elemzésnek ebben a szakaszában a kritérium az volt, hogy a zöngé a zár felpattanását megelőzően induljon.

A **VOT időtartamának** meghatározásához a zöngétlen megvalósulásokat a fonológiai zöngesség és a pontosság alapján kategóriába soroltuk (*p*, *t*, *k* és zöngétlenül ejtett *b*, *d*, *g*). Az időtartamra vonatkozó elemzést a fentiekben ismertetettek (életkor, nem, képzés helye, fonológiai zöngesség) túl kiegészítettük a **célszavankénti** megvalósulásokra. A zöngkezdesi idő meghatározásához a zárszakasz feloldásának kezdetét (többszörös felpattanás esetén az elsőét) és az abszolút zöngéindulást jelöltük.

A negatív VOT időtartamainak meghatározásához azokat a megvalósulásokat elemeztük, melyek alacsony frekvenciatartományban mutattak a spektrogramon energiakoncentrációt. Bár az előzöngé gyakorisága vizsgálatánál figyelembe vettük a különféle stratégiák (vö. Grácsi 2013) alkalmazásával létrehozott produkciókat, az időtartam vizsgálatánál törekedtünk az elemzést azokra a megvalósulásokra korlátozni, melyek esetében a zárt szájjüreg feltételei mellett jelenik meg a zöngé. A jelen vizsgálat elemzési módszere azonban nem alkalmas arra, hogy például az orrüreg megnyitásával létrehozott

szóalakokat teljes pontossággal kizárja, így a negatív VOT-re vonatkozó adatokat tájékoztató jellegűnek szántuk.

A hanganyag tartalmazott néhány olyan szóalakot, melynél a felvétel zajos volt, illetőleg a zárhang felpattanás nélkül realizálódott, vagy frázisban hangzott el. Ezeket a szóalakokat nem elemeztük.

A statisztikai elemzéshez az SPSS 20.0 szoftvert használtuk. Az adatokat a kiugró értékek kizárása után (mivel a normalitás feltételei továbbra sem teljesültek) nem parametrikus eljárásokkal elemeztük.

A vizsgálatban 58 tipikus nyelvi fejlődésű gyermek vett részt, három életkori csoportban (2. táblázat). A nemek aránya korcsoportonként közel azonos volt. A gyermekeket egy dél-dunántúli megyeszékhely három óvodájából választottuk ki.

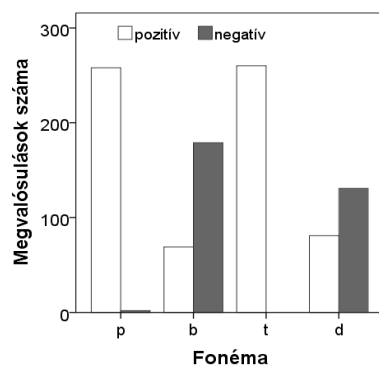
2. táblázat: A beszélők demográfiai adatai

Életkor: év;hó (átlag)	Résztevők		
	Összesen	Fiúk	Lányok
3;0–3;5 (3;3)	17	8	9
3;6–3;11 (3;8)	25	12	13
5;6–5;11 (5;8)	16	8	8
Összesen	58	28	30

Eredmények

Az előzőngés explozívák gyakorisága

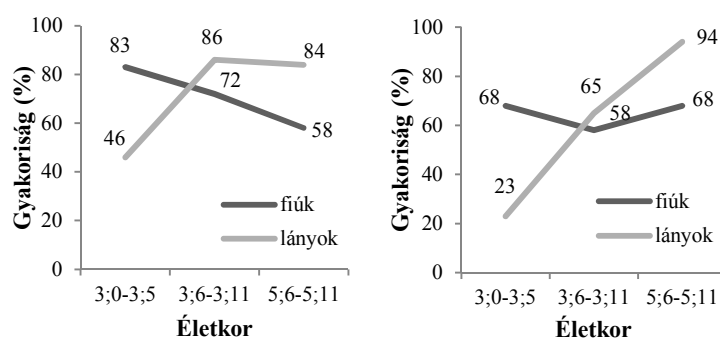
Elöl képzett zárhangok; 3;0–3;5, 3;6–3;11, 5;6–5;11 korcsoport. Az 1. ábra a negatív és pozitív VOT-értékek számát szemlélteti fonémánként a teljes mintára vonatkozóan. Eltekintve a p két előzőngés megvalósulásától, a zöngétlen hangok pozitív zöngékezdesi idővel realizálódtak, a zöngések VOT-i közt azonban negatív és pozitív értékek egyaránt megjelentek. A zöngés célfonémák megvalósulásait tovább elemeztük az életkor, a nem, a képzés helye és az előzőnge kapcsolatának feltárásához. A log-lineáris elemzésben a háromváltozós eredményezte a végső modellt [valószínűségi hányados: $\chi^2(10) = 2,81$, $p = 0,99$], az **életkor*nem*előzőnge interakció** szignifikáns hatásával. Az elemzést követő χ^2 -próba a lányok esetében szignifikáns kapcsolatot tárt fel az életkor és előzőnge közt [$\chi^2(2, N = 254) = 51,98$, $p < 0,001$], ami azt jelentette, hogy a lányoknál nőtt az életkor előrehaladtával az előzőngés megvalósulások száma. Továbbá a nem a legfiatalabb és a legidősebb vizsgált korcsoportban befolyásolta szignifikánsan az előzőnge gyakoriságát [$\chi^2(1, N = 143) = 25,45$, $p < 0,001$; $\chi^2(1, N = 134) = 12,50$, $p < 0,001$]. A nemenkénti eltérés iránya azonban a két életkori szakaszban eltérően alakult, míg a 3;0–3;5 életkorban a fiúk (77% vs. 36%), addig 5;6–5;11 éves korban a lányok (62% vs. 89%) ejtettek több előzőngés explozívát.



1. ábra

Az előzöngé gyakorisága fonémánként

Ahogy azt a 2. ábra mutatja, a fiatalabb korcsoportban tapasztalható jelentősebb nemenkénti eltérés a *d*-nek a lányoknál mért alacsonyabb számából következik leginkább.



2. ábra

Az előzöngé gyakorisága a *b* (balra) és *d* (jobbra) megvalósulásaiban

Veláris explozívák; 5;6–5;11 korcsoport. A *k* minden esetben pozitív, a *g* pozitív és negatív VOT-vel realizálódott. Az előzöngés veláris explozíva gyakorisága az elől képzettekre leírt tendenciának megfelelően alakul (fiúk: 78%, lányok: 96%), ami azt jelenti, hogy a veláris esetében is a lányok ejtettek több előzöngés hangot. A statisztikai elemzés a két nem adatai közt tendenciaszintű eltérést tárt fel [$\chi^2(1, N = 46) = 3,07, p = 0,08$].

A VOT időtartama

Pozitív VOT-k; elől képzett explozívák; 3;0–3;5, 3;6–3;11, 5;6–5;11 korcsoport. A 3. táblázat a zöngétlen realizációk (*p*, zöngétlenül ejtett *b*, *t* és zöngétlenül ejtett *d*) VOT-értékeinek korcsoportonkénti és nemenkénti leíró statisztikai adatait szemlélteti. A „zöngétlenül ejtett *d*, 3;0–3;5, fiúk” VOT kivételével a mediánok a rövid pozitív VOT kategóriájába esnek, az átlagos értékek többségében tehát aspirálatlan zárhangokat reprezentálnak. Az ábrából látható azonban, hogy számos esetben volt példa a magyar nyelvre nem jellemző hosszú zöngekezdesi időre is.

3. táblázat: Elöl képzett explozívák +VOT-értékei (időtartam ms-ban)

	<i>p</i>		Zöngétlenül ejtett <i>b</i>		<i>t</i>		Zöngétlenül ejtett <i>d</i>	
	Fiú	Lány	Fiú	Lány	Fiú	Lány	Fiú	Lány
3;0–3;5								
N	39	40	6	21	40	41	9	25
Átlag	22,1	27,3	26	25,5	25,0	27,0	34,6	20,6
Medián	18,0	20,5	16,5	18,0	20,0	24,0	35,5	19,0
Átlagos eltérés	11,1	17,9	23,6	17,3	18,0	14,9	7,9	8,6
Minimum	10	8	10	8	9	8	21	9
Maximum	54	79	61	79	70	73	43	39
3;6–3;11								
N	57	58	15	8	57	58	19	17
Átlag	22,2	21,7	29,1	21,7	25,0	22,5	20,0	21,2
Medián	18,0	18,0	20,0	19,5	22,0	19,5	21,0	19,0
Átlagos eltérés	13,2	13,0	19,3	14,1	14,3	11,2	8,2	8,2
Minimum	7	6	11	4	5	8	7	10
Maximum	59	62	75	44	82	60	34	36
5;6–5;11								
N	36	39	14	6	36	34	8	2
Átlag	21,6	20,6	21,5	12,8	28,2	17,1	17,3	12,5
Medián	18,5	21,0	16,5	12,5	24,5	15,0	17,5	12,5
Átlagos eltérés	10,9	11,4	11,7	3,6	16,9	6,8	5,4	6,3
Minimum	8	7	5	9	7	7	8	8
Maximum	52	51	47	19	70	37	25	17

A **korcsoportonkénti különbség** feltárására Kruskal–Wallis-tesztet használtunk, nemenként és fonémánként vizsgálva az adatokat. Az életkor két csoportban befolyásolta szignifikánsan a +VOT időtartamát: a fiúk csoportjában a zöngétlenül ejtett *d*, a lányoknál a *t* megvalósulásaiban [$H(2) = 10,31$, $p < 0,01$; $H(2) = 11,05$, $p < 0,01$]. Mindkét hang esetében az átlagos időtartam csökkent az életkorral.

Mann–Whitney-tesztel elemeztük az adatokat a **nemenkénti különbség** feltárásához, a Bonferroni-korrekción eredményeképp a hatás 0,016 szignifikanciaszinten kerül bemutatásra. Az elemzés alapján nemenkénti különbség a VOT időtartamában a következő csoportokban található: 3;0–3;5 éves korban a zöngétlenül ejtett *d* esetében, 5;6–5;11 éves korban a *t* esetében (U : 18, $z = -2,85$, $p < 0,01$; U : 400, $z = -2,49$, $p = 0,01$). Mindkét alkalommal a fiúk produkáltak hosszabb átlagos zöngékezdési időtartamot.

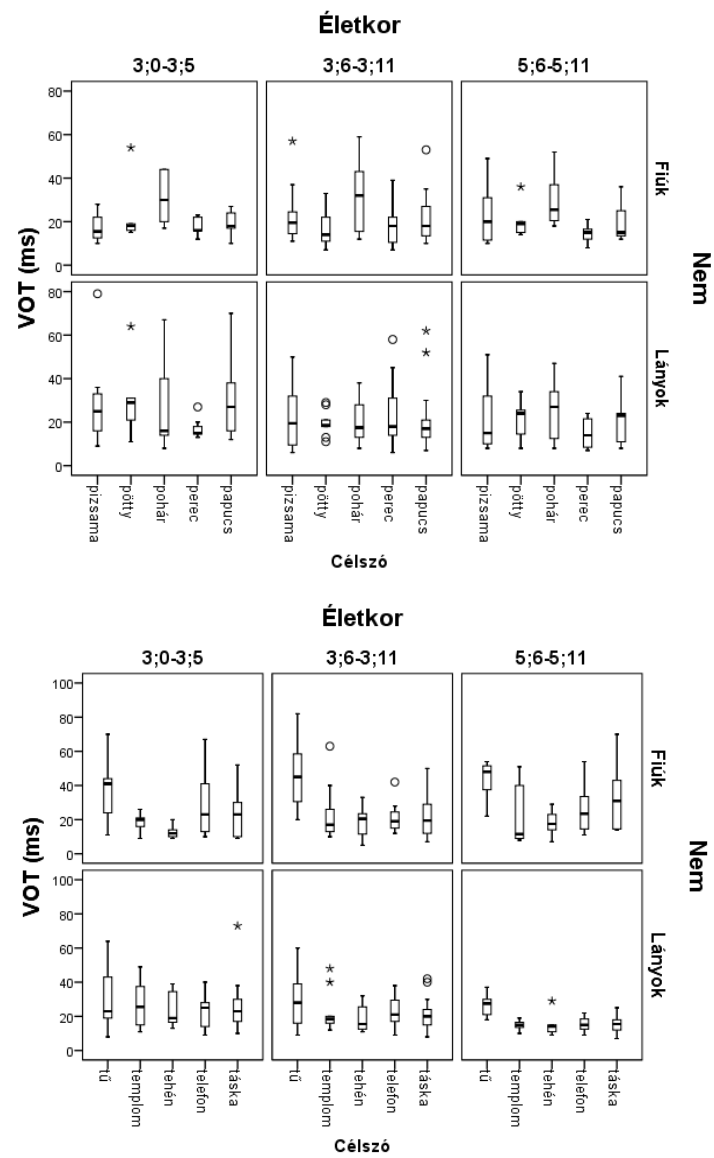
A **fonémakategóriák közti különbség** feltárásához az artikuláció helye (*p-t*; *b-d*) és a fonológiai zöngésség (*p-b*; *t-d*) tekintetében vizsgáltuk a párokat. A Wilcoxon-féle rangpróba nem tárt fel statisztikailag igazolható különbséget egyik vizsgált paraméter esetében sem.

A 3. ábra a *p* és *t* különböző célszavakban kapott VOT-értékeinek eloszlását jeleníti meg, az eredményeket korcsoportonként és nemenként elkülönítve ábrázoltuk. A **célszó hatását** a zöngétlenül ejtett zöngés célfonémák megvalósulásaira a minta kis elemszáma miatt nem vizsgáltuk. A Friedman-féle ANOVA-teszt a fiúk esetében tárt fel szignifikáns különbségeket. A *p*-re a célszó hatása tendenciaszintű 5;6–5;11 éves korban [$\chi^2(4) = 8,15$, $p = 0,086$], a *t*-re szignifikáns 3;0–3;5 és 3;6–3;11 éves korban [$\chi^2(4) = 11,26$, $p < 0,05$; $\chi^2(4) = 17,45$, $p < 0,01$]. A hosszú pozitív VOT jellemzően bizonyos szóalakokhoz kapcsolódott (pl. *táska*, *pohár*, *papucs*).

Pozitív VOT-k; veláris explozívák, 5;6–5;11 korcsoport. A *k*-ra és a zöngétlenül ejtett *g*-re kapott VOT-k nemenkénti statisztikai adatait a 4. táblázat tartalmazza. A zöngétlenül ejtett *g* alacsony elemszáma miatt további elemzéseket csak a *k*-ra vonatkozóan végeztünk. A Mann–Whitney-teszt alapján a *k* VOT-értékeiben a **nemek közti különbség** nem volt szignifikáns.

A *k*-ra kapott adatok eloszlása a 4. ábrán látható, együtt ábrázolva az ebben az életkorban mért (és a fentiekben már ismertetett) *p*-re és *t*-re kapott adatokkal. A **képzési hely hatása** a három vizsgált zárhang esetében szignifikáns [Friedman-teszt, fiúk: $\chi^2(2) = 25,01$, $p < 0,001$; lányok: $\chi^2(2) = 39,71$, $p < 0,001$], a veláris értékei jelentősen magasabbak az elől képzettekénél.

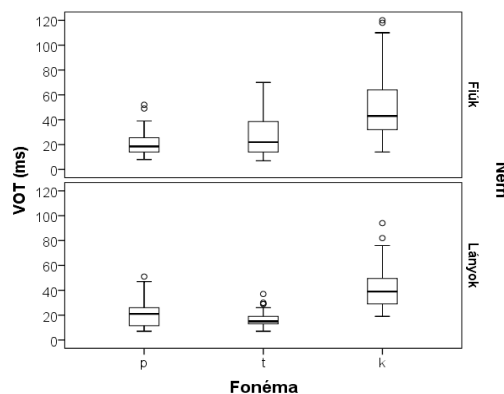
A **célszó hatása** a lányok mintájában volt feltárható [$\chi^2(7) = 11,54$, $p < 0,05$]. A szavak közti különbséget a legrövidebb mediánt (27,5 ms) és a legalacsonyabb átlagos eltérést (5,4 ms) mutató *karácsony* szó eredményezte. A szólista további szavai némiképp kiegyenlítettebb értékeket mutattak (medián, majd átlagos eltérés: *kutya*: 50,7 ms, 17,2 ms; *kard*: 40,5 ms, 13,4 ms; *kés*: 43,0 ms, 17,4 ms), kivéve a *könyv* szóban kapott értékeket (34,0 ms, 24,8 ms). Az eredmény, miszerint a *karácsony* szóban megvalósuló *k* VOT-értéke jelentősen rövidebb a többi (kivéve a *könyv*) szóban megvalósuló velárisokétól, vélhetően a szótagszámmal van összefüggésben; ebben a vonatkozásban érdekes eredmény, hogy a fiúk mintájában a szótagszám hatása nem jelentkezett.



3. ábra
A *p* (felül) és *t* (alul) megvalósulások célszavankénti VOT-értékei

4. táblázat: Velárisok VOT-értékei (időtartam ms-ban) a 5;6–5;11 korban

	<i>k</i>		Zöngétlenül ejtett <i>g</i>	
	Fiúk	Lányok	Fiúk	Lányok
N	37	39	5	1
Átlag	49,7	41,0	38,6	43,0
Medián	43,0	39,0	40,0	43,0
Átlagos eltérés	25,4	17,8	8,0	–
Minimum	14	19	25	–
Maximum	120	94	45	–



4. ábra

A *p*, *t*, *k* megvalósulások VOT-értékei

Bár a beszédminta kiváltásához használt szólista a célhang fonetikai környezetét illetően nem volt rendszerszerűen tervezve, a *könyv* és a *kés* szavakban kapott megvalósulások megengedik bizonyos hatások megfigyelését. Nevezetesen, a követő magánhangzó ajakállása tekintetében az eredmények megfelelnek Gósy (2000) eredményeinek, mely szerint a velárisok zöngékezési idejét a követő magánhangzó úgy befolyásolja, hogy labiálisok előtt rövidebbé válik a VOT értéke.

A negatív VOT-k sajátosságai; 5;6-5;11 korcsoport. A negatív VOT-val megvalósított explozívák közül csak azokat vizsgáltuk, melyek a spektrogramon ellenőrizve az alacsony frekvenciatartományban mutattak energiakonzentrációt. A két nem esetében közel azonos arányban találtunk a kritériumnak megfelelő megvalósulásokat az előzöngés realizáción belül (34/59, azaz 58% a fiúk, és 50/83, azaz 60% a lányok mintájában). Ez alapján úgy tűnik, hogy a lányok fentiekben bemutatott adatai, miszerint több előzöngés zárhangot ejtenek, mint a fiúk, abból adódik, hogy gyakrabban élnek a felnőtthöz eltérő stra-

tégiakkal, mint a fiúk. Az előzőngés realizációk további vizsgálatára van szükség e feltevés pontosításához, illetőleg megerősítéséhez.

Az elemzésben feltárt minimumértékek (p , $t \approx 7$ ms, $k = 14$ ms, valamint b , d , $g = -18$ ms) alapján a pontosan ejtett zöngés és zöngétlen hangok a VOT-dimenzió jól körülhatárolható tartományával elkülöníthetők.

A zöngés realizációk VOT-értékeinek nemenkénti statisztikai adatait az 5. táblázat tartalmazza. A mediánok közt a Mann–Whitney-teszt alkalmazásával szignifikáns **nemenkénti különbség** nem volt feltárható. Továbbá a Friedman-teszt eredménye alapján nem különbözött az időtartam a **képzési hely** vagy a **célszó** hatásaként sem. A képzés helye a felnőttek produkciójában sem befolyásolja szignifikánsan a negatív VOT időtartamát (Gósy–Ringen 2009).

5. táblázat: A zöngés megvalósulások VOT-értékei (időtartam ms-ban) az 5;6–5;11 korosztályban

	<i>b</i>		<i>d</i>		<i>g</i>	
	Fiúk	Lányok	Fiúk	Lányok	Fiúk	Lányok
N	7	20	14	20	13	10
Átlag	–73,2	–55,9	–53,7	–60,4	–67,7	–50,6
Medián	–69,0	–54,5	–54,5	–68,5	–61,0	–48,0
Átlagos eltérés	23,8	23,2	21,9	25,1	27,5	21,8

Következtetések

A jelen tanulmány célja az volt, hogy képmegnevezés során nyert szó eleji explozívák zöngésség szerint produkciójában a 3;0–3;5, 3;6–3;11 és 5;6–5;11 éves korban fellelhető sajátosságokat bemutassa. VOT-elemzéssel az előzőngés explozívák gyakoriságát, illetőleg a VOT-k időtartamát vizsgáltuk a tekintetben, hogyan változik a mintázat különböző nyelvi és nem nyelvi tényezők hatására.

Az **előzőngés explozívák gyakoriságában** a legfiatalabb és legidősebb korcsoportban nemenkénti, eltérő irányú eltérést találtunk. Első hipotézisünk, mely szerint az előzőnge gyakorisága 5;6–5;11 éves korcsoportban nem különbözik a fiúk és lányok mintájában, így nem igazolódott be.

A 3;0–3;5 korcsoport mintázata közti eltérés feltehetően fonológiai fejlődésbeli különbséget tükröz, mely szerint a fiúk előbb sajátítják el a szó eleji zöngésségi kontrasztot, mint a lányok. Hasonló eredményt kaptak vizsgálatukban Smit és munkatársai (1990) is. A jelenség magyarázata azonban további vizsgálatot igényel.

Az **5;6–5;11 éves korban** tapasztalt eltérés (az előzőngés elől képzett explozívák szignifikánsan, a velárisok tendenciaszerűen magasabb értéke a lányok mintájában) nem magyarázható fejlődési tényezőkkel; tekintve a fiatalabb korosztályra vonatkozó, fentebb bemutatott adatokat, valamint hogy vizsgálati eredmények szerint ebben az életkorban már a kontrasztrendszer (a

produktív tudást illetően is) teljesen kiépül (Sebestyén 2006). A **fonológiai fejlődésen kívül álló tényezők** magyarázhatják a jelenséget, hogy bizonyos beszédhelyzetekben a lányok nagyobb figyelmet fordítanak a gondosabb artikulációra (inkább produkálják a motorosan komplexebb, nagyobb erőfeszítéssel járó, de fonémikusan pontos előzöngés hangot), mint a fiúk (akik ugyanebben a helyzetben a motorosan kevésbé összetett, könnyebben létrehozható, de nyelvi funkciójában nem megfelelő hangot ejtik nagyobb gyakorisággal). Mátyus (2013) a morfológia területén tapasztalta, hogy 5 éves lányok szignifikánsan gyakrabban használják a bAn sztenderd változatát, mint az ugyanolyan korú fiúk (olyan szituációban, mikor a beszédpartner nyelvváltozata közelebb áll a sztenderdhez, mint az adatközlőé). Annak megválaszolására azonban, hogy valóban szociofonetikai tényezők állhatnak-e a feltárt nemek közti különbség hátterében, további vizsgálatokat igényel.

A **pozitív VOT-k időtartama**. A *p*-re és *t*-re a különböző életkori csoportokban kapott mediánokat a felnőttek értékeivel összevetve azt találjuk, hogy az időtartamok a gyermekek mintájában hosszabbak, különösen a *p* esetében. Kivételt képeznek az 5;6–5;11 éves lányok, akiknél a *t*-re mért értékek már felnőttzerűek, esetükben a zöngétlen alveoláris VOT-értéke az életkorral csökken. A *k*-ra 5;6–5;11 éves korban mért átlagos VOT a fiúk csoportjában lényegesebben, a lányokéban enyhén hosszabb a felnőttekre kapott adatoknál (vö. Gósy–Ringen 2009). Az eredmények alapján a gyermekek nagy része ún. éretlen kontrasztot (vö. Scobbie et al. 2000) használ a zöngésség jelölésében.

Nemekénti különbséget a VOT-k időtartamában a legfiatalabb és legidősebb vizsgált életkori csoportban és a zöngétlenül ejtett *d* és *t* esetében találtunk. Az eredmények megegyeznek azokkal a kutatási beszámolókkal, melyek a **nem, az életkor és az explozív típusa** interakcióját találták a VOT értékét befolyásoló tényezőnek (Bóna–Ausmann 2014; valamint Whiteside–Marshall 1998; Whiteside et al. 2004; Karlsson et al. 2004). A VOT-k időtartamát tekintve azt feltételeztük, hogy nemek közti eltérés – ha fennáll – csak bizonyos fonémák és bizonyos életkori csoport esetében lesz kimutatható. Feltételezésünk beigazolódott.

Fonémakategóriák közti különbséget az elől képzett explozívák közt nem igazoltunk. Tehát, eltérően a felnőtteknél tapasztalt, képzési helytől függően változó VOT-értékektől (Gósy–Ringen 2009), a (bilabiális és alveoláris) képzési hely hatása a jelen vizsgálatban nem befolyásolta a zöngékezdés időtartamát. Az eredmények azonban hasonlóak Bóna és Ausmann (2014) gyermekekre dokumentált eredményeihez, akik szintén nem találtak eltérést a *p* és *t* értékei közt a fiatalabb korosztályban, a képzési hely szerinti differenciálódás az elől képzett explozívákra vonatkozóan 11 éves kortól volt jellemző.

A jelen vizsgálat eredménye, miszerint a *p* és *t* VOT-értékei közti különbség nem jelentős, a *p*-re kapott hosszabb időtartamból adódhat. Figyelembe véve, hogy a bilabiálisra nézve sem az életkornak, sem a nemnek nem volt befolyásoló hatása, feltehető, hogy a felnőtteknél kapott adatoktól való elté-

rés nem (pusztán) módszertani, inkább fejlődési tényezőkkel magyarázható. A beszédszervi mozgás motoros szabályozásának kutatása kapcsán a bilabiális zár létrehozásával kapcsolatban jelentek meg felnőttek és gyermekek adatai közt olyan különbségek, melyek indokolhatják a zöngéindulás késését a bilabiális esetben (pl. állkapocsmozgás sebessége, vö. Smith 2010; ajkak és állkapocs eltérő részvétele a zár kialakításában, pl. Green et al. 2010).

A zöngétlen explozívák VOT-értékeinek képzési hely szerinti mintázata (az 5;6–5;11 korcsoportban mérve) a felnőttekre dokumentálthoz abban hasonlít, hogy veláris képzési helyen szignifikánsan (a leg)hosszabb a zöngékezdési időtartam (Gósy–Ringen 2009). A hosszabb időtartam a zárfelbontásban részt vevő artikulációs szerv ajkaknál vagy nyelvhegynél nagyobb tehetetlenségével, kevésbé mozgékony voltával lehet összefüggésben, melynek következtében a zöngéinduláshoz szükséges aerodinamikus feltételek később állnak elő, mint a mozgékonyabb elülső képzési helyen működő szervek esetében. A hatás felnőttek beszédében is meglehetősen robosztus; az elülső képzési helyeken tapasztaltakkal ellentétben például beszédstílustól (Neuberger 2014) és az időződés hatásától (Bóna 2011) függetlenül is megnyilvánul.

A fonológiai zöngéesség nem befolyásolta a zöngétlen realizációk időtartamát, mely eredmény azt jelzi, hogy fedett kontraszt használata nem jellemző a vizsgált gyermekek beszédprodukcijára. A jelenséget magyarázhatja a fenti elképzelés, miszerint a zöngétlenül realizált zöngés explozívák jelenléte a vizsgált csoportban nem magyarázható a fonológiai fejlettséggel.

A célszó hatása az elől képzett explozívák esetében a fiúk, a *k*-nál a lányok esetében volt kimutatható. Az eredmények nem minden esetben tulajdoníthatók a követő magánhangzó nyelválásfoka hatásának. A célhang tágabb fonetikai környezete módosíthatta a követő magánhangzó képzése nyomán előálló aerodinamikus hatásokat. Annak feltárása, hogy e szóalakok mely tulajdonsága járult hozzá a VOT tartamnövekedéséhez, további vizsgálatokat kíván.

A zöngétlen realizációknak a VOT időtartamán kívüli további (a vizsgálat során az elől képzett explozívák megvalósulásaiban megfigyelt, de jelen tanulmányban nem elemzett) sajátossága, hogy többszörös felpattanások esetén a gyenge felpattanást követi az erősebb intenzitású. E mintázat ellentétes a felnőttekre dokumentálttal (vö. Grácz–Kohári 2012), de 6;7 éves átlagos életkorú atipikus nyelvfejlődésű gyermekek beszédprodukcijára jellemző sajátosság (S. Tar 2013). A jelenség a levegőáramlás szabályozásának még nem felnőttzerű fejlettségi szintjét tükrözheti. A többszörös felpattanások rendszerszerű elemzése, valamint a VOT további akusztikai jegyeinek feltárása hasznos információval szolgálhat a zöngéességi kontraszt gyermeki produkciója megismerését illetően.

A negatív VOT sajátosságai az 5;6–5;11 éves életkori szakaszban. A mediánok közel abba a tartományba esnek, a maximum értékek azonban messze meghaladják azt az időtartamot (64 ms), melyet Ohala (1983) talált felnőttekkel végzett vizsgálatában a kizárólag passzív stratégia alkalmazásá-

val nyert zöngekezdési időre vonatkozóan. Ez alapján feltételezhető, hogy a vizsgálatban szereplő gyermekek valamilyen aktív stratégiával (pl. állkapocs ejtése) éltek az előzönge létrehozásakor. Az orrüreg felé való nyitással megvalósuló realizációkat próbáltuk kizárni azzal, hogy csak az alacsony frekvenciás VOT-ket elemeztük. Ahogy a bevezetőben is említettük, a jelen kutatásban alkalmazott eljárás azonban nem alkalmas a zönge spektrális tulajdonságainak feltárására (így a nazalitást mutató realizációk pontos felismerésére sem), e tekintetben további vizsgálatokra van szükség. A bemutatott eredményeket tájékoztató jellegűnek szántuk.

Összegezve, a jelen kutatás eredményei az életkor és a nem interakciójának hatását mutatták az előzönge gyakoriságára; hasonló interakciót tártunk fel a zöngétlen megvalósulások zöngekezdési ideje tekintetében is bizonyos fonémák, illetőleg bizonyos fonetikai környezet vonatkozásában. Amellett érveltünk, hogy az eredmények részben a beszédmotoros készségek fejlettségi szintjéből adódnak, valamint hogy a VOT észlelt mintázata nem minden esetben magyarázható fejlődési tényezőkkel. Végül, a negatív VOT-k több esetben jelenítettek meg magasabb frekvenciákon is energiakonzentrációt. A hangminőség megállapításához, ezzel együtt az előzönge képzéséhez alkalmazott stratégia feltárásához további vizsgálatokra van szükség. A minta kis elemszáma miatt az eredmények nem általánosíthatók.

Irodalom

- Boersma, Paul – Weenink, David 2011. *Praat: doing phonetics by computer*. [Software]. 5.3. verzió. <http://www.praat.org/> (A letöltés ideje: 2011. október 10.)
- Bóna Judit 2011. A [p, t, k] mássalhangzók zöngekezdési ideje idők és fiatalok spontán beszédében és felolvasásában. *Beszédkutatás* 2011. 61–73.
- Bóna Judit – Auszmann Anita 2014. Voice onset time in language acquisition: Data from Hungarian. In: *Proceedings of the 10th International Seminar on Speech Production*. Cologne. 41–44.
- Dodd, Barbara – Holm, Alison – Hua, Zhu – Crosbie, Sharon 2003. Phonological development: A normative study of British English-speaking children. *Clinical Linguistics and Phonetics* 17/8. 617–643.
- Gandour, Jack – Petty, Soranee H. – Dardarananda, Rochana – Dechongkit, Sumalee – Mukongoen, Sunee 1986. The acquisition of the voicing contrast in Thai: A study of voice onset time in word-initial stop consonants. *Journal of Child Language* 13. 561–572.
- Gósy Mária 1984. *Hangtani és szótani vizsgálatok hároméves gyermekek nyelvében*. Nyelvtudományi Értekezések 119. Akadémiai Kiadó, Budapest.
- Gósy Mária 1998. A szavak hangalakjának változása a gyermeknyelvben. *Beszédkutatás* '98. 1–39.
- Gósy Mária 2000. A [p, t, k] mássalhangzók zöngekezdési ideje. *Magyar Nyelvőr* 124. 195–204.

- Gósy, Mária 2001. The voice onset time of the Hungarian voiceless plosives in words and in spontaneous speech. *International Journal of Speech Technology* 3–4. 155–164.
- Gósy Mária 2004. *Fonetika, a beszéd tudománya*. Osiris, Budapest.
- Gósy, Mária – Ringen, Catherine O. 2009. *Everything you always wanted to know about VOT in Hungarian*. Előadás az ICSH 2009 kongresszuson, Debrecen, 2009 szeptember 1. http://icsh9.unideb.hu/pph/handout/Ringen_Gosy_handout.pdf (A letöltés ideje: 2012. április 6.)
- Grácz Tekla Etelka 2012. *Zöreghangok akusztikai fonetikai vizsgálata a zöngésségi oppozíció függvényében*. PhD-értekezés. ELTE, Budapest.
- Grácz Tekla Etelka 2013. Explozívák és affrikáták zöngésségének időviszonyai. *Beszéd kutatás 2013*. 94–121.
- Grácz Tekla Etelka – Kohári Anna 2012. A zöngékezdesi idő egy módszertani kérdés függvényében. In: Markó Alexandra (szerk.): *Beszédtudomány. Az anyanyelv-elsajátítástól a zöngékezdesi időig*. ELTE BTK – MTA Nyelvtudományi Intézet, Budapest. 228–248.
- Green, Jordan R. – Moore, Christopher A. – Higashikawa, Masahiko – Steewe, Roger W. 2010. The physiologic development of speech motor control: Lip and jaw coordination. *Journal of Speech, Language, and Hearing Research* 43. 239–255.
- Karlsson, Fredrik – Zetterholm, Elisabeth – Sullivan, Kirk P. H. 2004. Development of a gender difference in voice onset time. In: *Proceedings of the 10th Australian International Conference of Speech Science and Technology*, Sydney. 316–321.
- Kewley-Port, Diane – Preston, Malcolm S. 1974. Early apical stop production: A voice onset time analysis. *Journal of Phonetics* 2. 195–210.
- Koenig, Laura L. 2000. Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds. *Journal of Speech, Language and Hearing Research* 43. 1211–1228.
- Lisker, Leigh – Abramson, Arthur S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20. 384–422.
- Mátyus, Kinga 2013. The (bAn) variable in the speech of five- and ten-year-old Hungarian children. In Polczyńska, Monika – Pakula, Lukasz P. – Jaworska, Dorota (eds.): *Young linguists' insights: Taking interdisciplinary approaches to the fore*. Wydział Anglistyki UAM, Poznań. 55–63.
- S. Meggyes Klára 1971. *Egy kétéves gyermek nyelvi rendszere*. Nyelvtudományi Értekezések 73. Akadémiai Kiadó, Budapest.
- Neuberger Tilda – Grácz Tekla Etelka 2013. Az alveoláris zöngétlen explozíva variabilitása. *Beszéd kutatás 2013*. 160–173.
- Neuberger Tilda 2014. Zöngétlen explozívak időviszonyai a beszéd típus és az egyéni sajátosságok tükrében. *Beszéd kutatás 2014*. 56–69.
- Nissen, Shown L. – Fox, Robert A. 2009. Acoustic and spectral patterns in young children's stop consonant productions. *Journal of the Acoustical Society of America* 126/3. 1369–1378.
- Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In MacNeilage, Peter F. (ed.): *The production of speech*. Springer-Verlag, New York. 189–216.

- Ohala, John J. 2011. Accomodation to the aerodynamic voicing constrant and its phonological relevance. In Lee, Wai-Sum – Zee, Eric (eds.): *Proceedings of the XVIIth International Congress of Phonetic Science*. Hong Kong. 64–67.
- Sebestyén Tar Éva 2006. *A 3–6 éves kori fonológiai fejlődés kronológiai mintázata a magyarban*. Open Art, Budapest.
- Sebestyén Tar Éva 2008. *Az atipikus nyelvi fejlődés szegmentális fonológiai szintjének elemzése*. PhD-értekezés. PTE, Pécs.
- Scobbie, James M. – Gibbon, Fiona – Hardcastle, William J. – Fletcher, Paul 2000. Covert contrast as a stage in the acquisition of phonetics and phonology. In Broe, Michael B. – Pierrehumbert, Janet B. (eds.): *Papers in Laboratory phonology V: Acquisition and the lexicon*. Cambridge University Press, London. 192–203.
- Siptár, Péter – Törkenczy, Miklós 2007. *The phonology of Hungarian*. Oxford University Press, Oxford.
- Smit, Ann B. – Hand, Linda – Feilinger, Joseph J. – Bernthal, John E. – Bird, Ann 1990. The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorder* 55. 779–798.
- Smith, Anne 2010. Development of neural control of orofacial movements for speech. In Hardcastle, William J. – Laver, John – Gibbon, Fiona E. (eds.): *Handbook of phonetic sciences*. Blackwell, Oxford.
- So, Lydia K. H. 2006. Cantonese phonological development: Normal and disordered. In Zhu Hua – Barbara Dodd (eds.): *Phonological development and disorders in children: A multilingual perspective*. Multilingual Matters, Canada. 109–135.
- S. Tar Éva 2013. A zöngésségi kontraszt elsajátítottságának mintázata atipikus nyelvi fejlődés esetén. *Beszéd kutatás 2013*. 194–210.
- Whiteside, Sandra P. – Marshall, John 1998. Voice onset time pattern in 7-, 9- and 11-year old children. In Mannell, Robert H. – Robert-Ribes, Jordi (eds.): *Proceedings of the 5th International Conference on Spoken Language Processing*. Volume 6. Australian Speech Science and Technology Association. 2687–2690.
- Whiteside, Sandra P. – Henry, Luisa – Dobbin, Rachel 2004. Sex differences in voice onset time: A developmental study of phonetic context effects in British English. *Journal of the Acoustical Society of America* 116/2. 1179–1183.

A MAGÁNHANGZÓK AKUSZTIKAI SZERKEZETE 9 ÉS 11 ÉVES ISKOLÁSOK SPONTÁN BESZÉDÉBEN

Auszmann Anita

Bevezetés

A gyermekek születésük utáni első éveik során fokozatosan sajátítják el az anyanyelvüket. Beszédükben először mennyiségi fejlődés, majd 6 éves kor után finom minőségi változások figyelhetők meg (Gósy 2005). A változás mértéke a 6 éves kor előtti periódusban igen jelentős, éppen ezért került a kutatások középpontjába. Korántsem zárul le azonban ennek a korszaknak a végére. Sokat kutatták azt a szakaszt is, amikor a gyermekek belépnek az iskolába, hiszen az intézményes oktatás szintén jelentős hatással van a beszédre (Neuberger 2014).

A gyermekek 9 éves korra már két-három évet eltöltöttek az oktatásban, ennek hatása megfigyelhető beszédükben; az alsó tagozatból a felső tagozatba lépés azonban a nyelvi fejlődés szempontjából szintén érzékeny periódus lehet fejlődésükben. A szakirodalom alapján ismert, hogy a beszédhangok differenciálása még kilencéves korban sem tökéletes, sőt az adatokból némi fejlődési megtorpanásra következtethetünk a magyarban (Gósy 2006). Az angol nyelvben végzett vizsgálatok azt mutatták, hogy 9 és 12 éves kor között a beszédhangok időtartamának tartománya és változatossága az életkor előrehaladtával szignifikánsan csökken (Lee et al. 1999). A zárhangok zöngékezdési időértékei 9 éves korban még jóval nagyobb szórást mutatnak, mint két évvel később (Whiteside et al. 2003). 9 éves korra eltűnik az az alaphangmagasságban tapasztalt különbség, ami megfigyelhető fiúk és lányok között 7 éves korban (a fiúk alaphangmagassága még magasabb, mint a lányoké), ami nem változik 11 éves korra sem, vagyis fiúk és lányok alaphangmagassága nagymértékben hasonló marad (Auszmann–Neuberger 2014). Az életkor előrehaladtával a beszédtempó folyamatos gyorsulást mutat (Gósy 2003; Gyarmathy 2007; Laczkó 2009), a gyermekek spontán beszédében csökken az agrammatikus mondatok aránya (Laczkó 2011), egyre komplexebb szerkezeteket hoznak létre, gyakoriak az összetett mondatok, közléseikben sok és változatos kötőszókat használnak (Horváth 2013). Ekkor még jóval nagyobb arányban jelennek meg szünetek a gyermekek beszédében, mint a felnőtteknél. Fiúk és lányok között még nagy különbség van a narratívájukat illetően, a lányok szívesebben beszélnek, több szóból álló beszédszakaszokat hangosítanak meg, néma szüneteik rövidebbek. 9 éves korra minden nyelvtani szerkezetet és szabályt helyesen tudnak alkalmazni (Neuberger 2011), de még nem ala-

kulnak ki a pragmatikai szabályok (Horváth 2013). Mindezen vizsgálatok eredményeit, illetve a 9 éves kor utáni nyelvfejlődésről szerzett ismereteinket egészíti ki, ha képet kapunk arról, hogyan stabilizálódik a magánhangzók képzése és akusztikai szerkezete ezen életkorban, amely alapján feltérképezhető a hangzókészlet elsajátítása mint az anyanyelv-elsajátítási folyamatok egyik meghatározó részfolyamata.

Az egyes magánhangzókat az első két formáns (F_1 , F_2) (a zöngének a rezonátorüregekben felerősödött felharmonikusa) egyértelműen meghatározza (Gósy 2004): az F_1 értéke az állkapocs nyitásszögével, a nyelvállás fokával van összefüggésben, míg az F_2 értéke a nyelv vízszintes mozgásával és az ajakműködéssel. A magánhangzótér a nemzetközi szakirodalomban a magánhangzók F_1 és F_2 koordinátái mentén ábrázolt vokális térként jelenik meg, vagyis a nyelv mozgása szerint egymástól legtávolabb eső hangok határozzák meg ezt a teret (vö. IPA Handbook 1999; Vorperian 2007). Az első és a második formáns értéke továbbá függ a toldalékcso hosszától, alakjától és térfogatától. A toldalékcso felnőttkorig jelentős változásokon megy keresztül, és ez hatással van a beszéd akusztikumára is. Ennek megfelelően a magánhangzók akusztikai szerkezetét befolyásolja a testi fejlettségi fok, amely bizonyos mértékben korrelál az életkorral, illetve a beszélő nemével (Fant 1966; Huber et al. 1999; Whiteside–Hodgson 2000; Perry et al. 2001; Deme 2012). A fiúknál és lányoknál különböző mértékű és ütemű a testméret-növekedés, illetve eltérő hormonális változásokon mennek keresztül. Nemzetközi kutatások eredményei azt mutatják, hogy már 4 éves korban különbség adatolható fiúk és lányok magánhangzóformánsainak frekvenciaértékei között (Perry et al. 2001), a fiúk F_1 és F_2 frekvenciaértékei minden életkorban alacsonyabbak, mint a lányoké (Lee et al. 1999), ami legmarkánsabban 12–15 éves kor között figyelhető meg.

Mind a nemzetközi, mind a hazai szakirodalomban vizsgálták már a magánhangzók frekvenciaszerkezetét és időtartamát tekintettel a korra, a nemre és a beszéd típusára. 6–7 éves óvodásokkal végzett vizsgálatok (Deme 2012) alapján elmondható, hogy spontán beszédben a gyermekek magánhangzóinak formánsai magasabb frekvenciaértékeken realizálódnak, mint a felnőttek ejtésében. Az egyes magánhangzókat vizsgálva az a megállapítás tehető, hogy a formánsértékek nagyobb szórással adatolhatók a gyermekeknél, mint a felnőttek ejtésében. Deme (2012) kutatásában nem volt szignifikáns különbség a gyermekek magánhangzóinak formánsértékeiben a nem tekintetében. 7 és 9 éves iskolás gyermekek spontán beszédében mért adatok (Auszmann 2014) azt mutatták, hogy a 7 éves fiúk és lányok magánhangzóinak akusztikai-fonetikai szerkezete nagyfokú variabilitást mutat a csoporton belül. 9 éves korra ez a különbség már csökken, a fiúk és a lányok adatai nagymértékben hasonlóvá válnak. Bár szignifikáns különbség adatolható a 7 éves és 9 éves gyermekek magánhangzói között, a formánsok frekvenciaértékeinek átlagai nem mutatnak nagymértékű különbséget. Az adatokat a felnőttek spontán be-

szédében mérhető adatokkal összevetve általánosságban megfogalmazható az, hogy az iskolások hangképzése stabilabb, mint az óvodásoké, de kevésbé stabil, mint a felnőtteké. Angol nyelvű gyermekekkel végzett kutatás szerint 9 és 12 éves kor között a beszédhangok időtartamának tartománya és változottsága csökken, 12 éves kor körül már a felnőttek beszédében mért értékekhez hasonlít (Lee et al. 1999).

A 9 éves korosztály spontán beszédében Bóna és Imre (2010) a fonológiaiilag rövid vs. hosszú magánhangzók egyértelmű elkülönülését adatolta a fizikai időtartamok alapján, csökken az ejtésbeli variabilitás. 9 évesek beszédében a fizikai időtartamok egyértelműen definiálják a fonológiaiilag rövid és hosszú magánhangzókat (Gósy 2006). A formánsszerkezet tekintetében kisiskolásoknál csak az F_2 esetében találtak szignifikáns különbséget a rövid és a hosszú hangok között (Bóna–Imre 2010). A spontán beszéd egyik jellegzetessége, hogy a beszédhangok nemegyszer indokolatlanul, illetve szándéktalanul megnyúlnak vagy megrövidülnek. Ezeknek az előfordulása beszélőnként változik, de kifejezetten jellemző a gyermekekre (Deme 2012). 7 és 9 éves gyermekek magánhangzóinak időtartamait összevetve azt találták (Auszmann 2014), hogy a legtöbb magánhangzó esetében szignifikáns rövidülés figyelhető meg az életkor előrehaladtával.

Egy adott gyermeknyelvi időszak magánhangzóiról a felnőtt adatok ismeretében lehet pontosabb megállapításokat tenni. Felnőttek (22–54 év) spontán beszédén végzett vizsgálatok azt mutatták, hogy a magánhangzók első és második formánsainak frekvenciaértékei meglehetősen nagy átfedést mutatnak. Jellemző tehát a magánhangzók semlegesedése, az alulkonfiguráltság következményeképpen pedig az, hogy a magánhangzó-realizációk kevésbé különülnek el egymástól (Grácsi–Horváth 2010). Fiatal beszélőkkel végzett kutatások igazolták, hogy a spontán beszédben a fizikai időtartam-különbség megvan a fonológiaiilag rövid-hosszú magánhangzópárok mindegyikénél (Gósy–Beke 2010). Fiatal és idős (70 év feletti) nők beszédét vetette össze Bóna (2009), és azt találta, hogy az idősök kisebb frekvenciatartományban képezik a magánhangzókat, mint a fiatalabbak, esetükben kevésbé különülnek el az ejtés során a különböző magánhangzó-minőségek. Azt is igazolta, hogy az idősök szignifikánsan hosszabb magánhangzókat ejtenek, mint a fiatalok; valamint hogy az életkor előrehaladtával a rövid és a hosszú hangok spektrális szerkezete kevésbé különül el, mint a fiatal beszélőknél. Olaszky (2006) és Bata (2007) megállapították, hogy a magánhangzó-időtartamokra nincs hatással a beszéd típusa, tehát mind spontán beszédben, mind felolvasásban hasonló értékeket mutatnak.

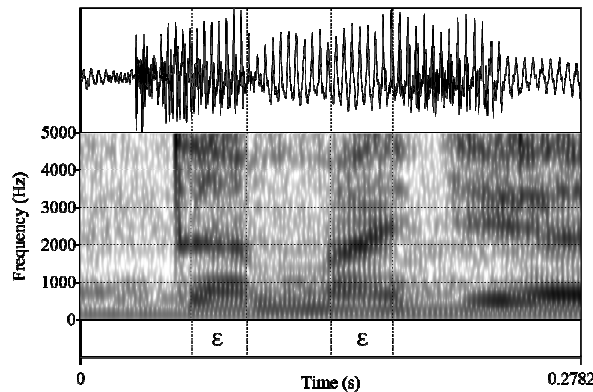
Jelen kutatásunkban arra kerestük a választ, hogy milyen eltérések adathatók a 9 és a 11 éves gyermekek spontán beszédében előforduló magánhangzók formánsszerkezete és időtartamértékei között. Mivel a korábbi, 9 évesekkel végzett kutatásunk szerint ebben az életkorban a magánhangzók frekvenciaszerkezete még a fiatalabb (7 éves) gyermekekéhez hasonlít job-

ban, így választ kerestünk arra is, hogy 11 éves korra milyen mértékben válik hasonlóvá a magánhangzók akusztikai-fonetikai szerkezete a felnőttekéhez. Hipotéziseink a következők voltak: a 11 éves gyermekek formánsértékei (i) nagyobb magánhangzótérben realizálódnak, (ii) az egyes magánhangzók frekvenciaértékei jobban elkülönülnek egymástól, mint a 9 éves gyermekekéi. (iii) A magánhangzók objektív időtartamértékei rövidülést mutatnak az életkor előrehaladtával. (iv) A lányok és fiúk magánhangzóinak formánsszerkezete között jelentős különbséget találunk mindkét életkori csoportban.

Anyag, módszer, kísérleti személyek

Kutatásunkban 9 és 11 éves tipikus fejlődésű, ép halló és ép intellektusú gyermekek vettek részt. Mindkét életkori csoportban 5 fiú és 5 lány szerepelt az adatközlők között. A gyermekekkel egyénileg, megszokott iskolai környezetben – lehetőség szerint zajmentes helyiségben –, tanítási időben készítettünk spontánbeszéd-felvételeket. A gyermekek az iskolájukról, a kedvenc játékaikról, illetve a szabadidős tevékenységükről beszéltek.

Mindegyik gyermek hanganyagából 1-1 percet elemeztünk (az átlagosan 4 perces felvételek közepéből). Kutatásunkban 9 magánhangzót vizsgáltunk fonetikai helyzettől és hangkörnyezettől függetlenül. Nem szerepelt külön vizsgálati szempontként a fonológiai rövid és hosszú magánhangzók összevetése. A Praat 5.3 (Boersma–Weenink 2011) programmal annotáltuk a magánhangzókat (összesen: 4571 db; 1. ábra), majd egy erre a célra készült szkripttel automatikusan kinyertük az egyes hangok első és második formánsainak frekvenciaértékét, valamint az időtartamát.



1. ábra

Példa a Praatban végzett szegmentálásra

Az annotálást manuálisan, a magánhangzók tiszta fázisának formánsstruktúrája, a második formáns kezdete és vége alapján végeztük. A szkript a fel-

címkézett magánhangzó tiszta fázisában mérte ki az első két formáns frekvenciaértékét a Praat alapértelmezett beállításai mentén (Burg-metóduson alapuló eljárással, 25 ms-os Hamming-típusú ablakkal, 5 formáns 5,5 kHz-ig) minden 10 ms-ban, majd a kapott értéknek vette a középértékét. A mért adatokat manuálisan ellenőriztük a mérési hibák elkerülése érdekében.

A kutatáshoz használt korpuszban szereplő magánhangzók darabszámát az 1. táblázat összegzi. A magánhangzók előfordulásának aránya megegyezik a felnőttek spontán beszédében megfigyelt arányokkal (Gósy–Beke 2010). A leggyakrabban az [e] és az [o] hang, míg legritkábban az [y] hang szerepelt a gyermekek spontán beszédében.

A kapott adatokon statisztikai elemzést (egytényezős ANOVA, Tukey-féle post hoc teszt) az SPSS 20.0 szoftverrel végeztünk 95%-os konfidencia-intervallumon.

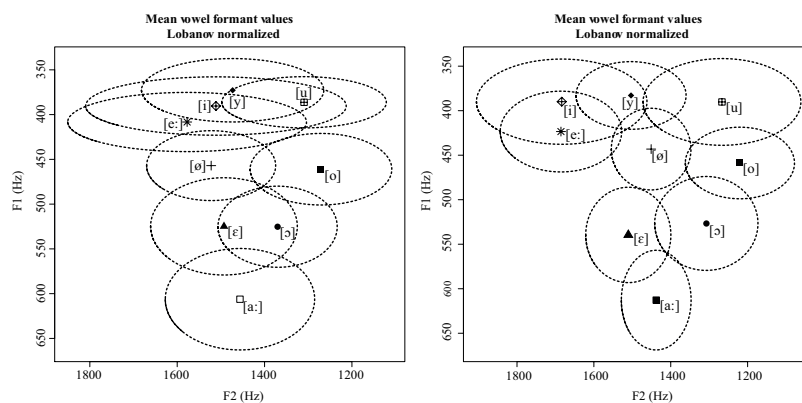
1. táblázat: A jelen kutatásban szereplő 20 gyermek spontán beszédében vizsgált magánhangzók darabszáma és aránya

V	[e]	[o]	[o]	[i]	[a:]	[e:]	[ø]	[u]	[y]
db	1190	978	675	594	418	339	165	146	66
%	26	21	15	13	9	8	4	3	1

Eredmények

A 9 és 11 éves gyermekek magánhangzóinak formánsszerkezete

A 9 és 11 évesek magánhangzóinak életkorok szerint normalizált formánsértékeit sűrűsödési ellipszisen ábrázoltuk (2. ábra).



2. ábra

A magánhangzók formánsértékei a 9 éves (balra) és 11 éves gyermekek (jobbra) spontán beszédében

A sűrűsödési ellipszisek jól mutatják, hogy 9 éves korban az egyes magánhangzók nagyobb területen realizálódnak, és nagyobb átfedések figyelhetők meg közöttük, mint 11 éves korban. Ez a különbség részben statisztikailag is igazolható (2. táblázat). Az F_1 tekintetében a legtöbb magánhangzó esetében szignifikáns különbség van az egyes életkorok között, csak az [e:] és [u] hangoknál nem. Ezzel szemben az F_2 esetében nagyrészt nem szignifikáns a különbség a két életkori csoport között, kivéve az [ɛ], [e:], [i] és [o] hangok esetében.

2. táblázat: A 9 és 11 éves gyermekek magánhangzóinak F_1 - és F_2 -paraméterei között adatolható szignifikáns különbségek

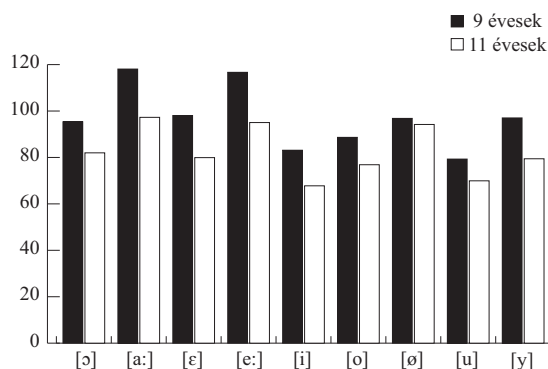
V	F ₁ -re kapott adatok			F ₂ -re kapott adatok		
	F	df	p	F	df	p
[ɔ]	23,594	977	0,001	2,506	977	0,058
[a:]	7,911	417	0,001	2,018	417	0,111
[e]	10,849	1189	0,001	10,533	1189	0,001
[e:]	1,493	338	0,216	8,978	338	0,001
[i]	5,437	593	0,001	30,421	593	0,001
[o]	12,687	674	0,001	3,079	674	0,027
[ø]	8,231	164	0,001	0,584	164	0,627
[u]	0,368	145	0,776	0,293	145	0,830
[y]	4,423	65	0,007	0,457	65	0,714

A 9 és 11 éves gyermekek magánhangzóinak időtartama

Megvizsgáltuk az egyes magánhangzók időtartamát (3. táblázat és 3. ábra). Az adatokból egyértelműen látszik, hogy a kor előrehaladtával csökken a magánhangzók átlagos időtartama. A megfigyeléseket a statisztikai elemzés is alátámasztja, a 9 és 11 éves gyermekek magánhangzóinak időtartama között csupán az [ø] és [y] hangok esetében nem adatoltunk szignifikáns különbséget (3. táblázat).

3. táblázat: A 9 és 11 éves gyermekek magánhangzó-időtartamai között talált szignifikáns különbségek

V	[ɔ]	[a:]	[e]	[e:]	[i]	[o]	[ø]	[u]	[y]
F	4,905	8,809	17,069	4,926	9,973	7,148	0,919	4,890	1,831
df	977	417	1189	338	593	674	164	145	65
p	0,002	0,001	0,001	0,002	0,001	0,001	0,433	0,003	0,151

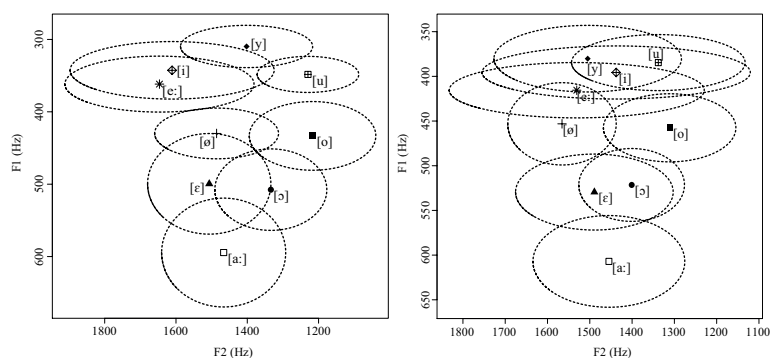


3. ábra

9 és 11 éves gyermekek magánhangzóinak átlagos időtartama

Nemek közötti különbségek

Megvizsgáltuk, hogy van-e, és ha igen, milyen mértékű különbség az egyes életkorokban a fiúk és a lányok magánhangzóinak között. Grafikonon ábrázoltuk a 9 éves fiúk és lányok (4. ábra) formánsainak frekvenciaértékeit. Annak ellenére, hogy minden életkorban a lányoktól várunk jobb teljesítményt, a sűrűsödési ellipszisek alapján az mondható, hogy 9 éves korban bár a fiúk magánhangzóinak is mutatnak átfedéseket, ezeknek mértéke az általunk vizsgált lányok esetében nagyobb. A legtöbb magánhangzó esetében nem adatoltunk szignifikáns különbséget a két nem között. A Tukey-féle post hoc teszt alapján 9 éveseknél szignifikáns különbség van fiúk és lányok között az F_1 esetében az [ɔ] ($p = 0,013$) hangnál, F_2 esetében pedig az [ɛ] ($p = 0,007$) és [i] ($p < 0,001$) hangoknál.



4. ábra

A 9 éves fiúk (balra) és lányok (jobbra) magánhangzóinak formánsszerkezete

A 4. táblázat összegzi a két nem spontán beszédében mért átlagos F_1 - és F_2 -értékeket, valamint magánhangzó-időtartamokat.

Nem figyelhető meg egyértelmű tendencia annak tekintetében, hogy mely magánhangzónál magasabbak a fiúk vagy a lányok F_1 -, illetve F_2 -átlagértékei. A legtöbb esetben a két nem átlagértékei közel azonosak, azonban az F_2 esetében, ahol jelentősebb különbség van a fiúk és a lányok átlagértékei között, ott rendszerint a fiúk értékei magasabbak. Az időtartamátlagokkal kapcsolatban megfigyelhető, hogy a lányok spontán beszédében valamivel hosszabban valósul meg a legtöbb magánhangzó.

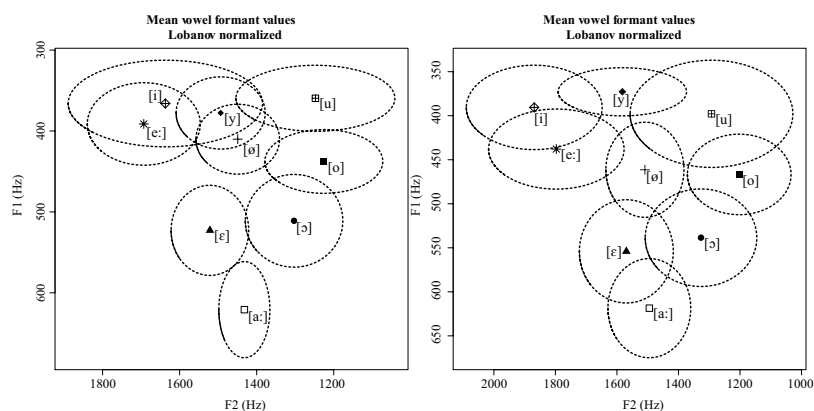
4. táblázat: 9 éves fiúk és lányok magánhangzóinak formánsértékei és időtartama (átlag és átlagos eltérés)

V	F_1 (Hz)		F_2 (Hz)		Időtartam (ms)	
	Fiúk	Lányok	Fiúk	Lányok	Fiúk	Lányok
[ɔ]	766±88	737±71	1631±222	1603±232	85±29	88±33
[a]	903±117	888±86	1823±282	1723±315	107±31	115±39
[e]	757±109	753±76	1909±299	1767±377	88±29	93±32
[ɛ]	547±67	557±59	2124±471	1836±570	104±37	112±39
[i]	518±62	523±51	2078±492	1694±610	77±25	78±26
[o]	653±73	634±66	1428±287	1476±231	79±25	87±28
[ø]	654±59	627±67	1898±194	1897±216	90±27	94±30
[u]	522±45	505±55	1444±244	1496±375	67±21	90±35
[y]	470±56	501±76	1887±262	1823±399	97±25	89±37

Ezt követően összevetettük a 11 éves fiúk és lányok (5. ábra) formánsainak frekvenciaértékeit is. Ebben az életkorban szintén azt találtuk, hogy a kutatásban részt vevő fiúk magánhangzói jobban elkülönülnek egymástól, mint a lányokéi, ugyanakkor ez a különbség nem mondható jelentősnek. A legtöbb magánhangzó esetében ebben az életkorban sem adatoltunk szignifikáns különbséget a két nem között. A Tukey-féle post hoc teszt alapján a 11 éveseknél statisztikai különbséget találtunk a két nem között az F_1 esetében az [ɛ] ($p = 0,01$), [ø] ($p = 0,018$) és [y] ($p = 0,009$) hangoknál, valamint az F_2 esetében az [i] ($p = 0,028$) hangnál. A 11 éveseknél mért F_1 -, F_2 - és időtartam adatoknak is kiszámoltuk az átlagait mindkét nem esetében, amit táblázatba foglaltunk (5. táblázat).

11 éves életkorban már az F_1 -nél és F_2 -nél is a legtöbb magánhangzónál a lányok esetében magasabbak az átlagértékek. A leggyakoribb magánhangzókat hasonló időtartamban valósították meg a fiúk és lányok is. A két életkor között két tendenciaszerű változás figyelhető meg az átlagértékek alapján: egyrészt az F_1 csökkenése mind a fiúknál, mind a lányoknál, másrészt a ma-

gánhangzó-időtartamok csökkenése szintén mindkét nem esetében. A statisztikai eredmények alapján az is elmondható, hogy sem a 9, sem a 11 éves korosztályban nincs különbség a magánhangzók időtartamában a két nem között. Az F_1 - és F_2 -paraméterek esetében is csak bizonyos magánhangzóknál, így a két nem között adatolható különbségek még nem tekinthetők jelentősnek ezekben az életkorokban.



5. ábra

A 11 éves fiúk (balra) és lányok (jobbra) magánhangzóinak formánsértékei spontán beszédben

5. táblázat: 11 éves fiúk és lányok magánhangzóinak formánsértékei és időtartama

V	F_1 (Hz)		F_2 (Hz)		időtartam (ms)	
	Fiúk	Lányok	Fiúk	Lányok	Fiúk	Lányok
[ɔ]	690±79	717±97	1551±214	1558±275	73±26	72±29
[a:]	831±94	853±96	1759±126	1810±206	94±27	90±28
[e]	710±83	744±97	1937±189	1926±239	75±26	75±24
[e:]	525±71	553±79	2229±282	2272±344	89±30	83±29
[i]	497±72	477±74	2170±455	2379±337	62±21	65±26
[o]	589±62	594±85	1397±301	1368±259	78±33	72±24
[ø]	548±68	597±109	1784±199	1816±212	89±38	81±26
[u]	487±59	493±95	1481±358	1513±399	59±20	71±25
[y]	507±70	438±40	1866±214	1902±308	64±20	75±23

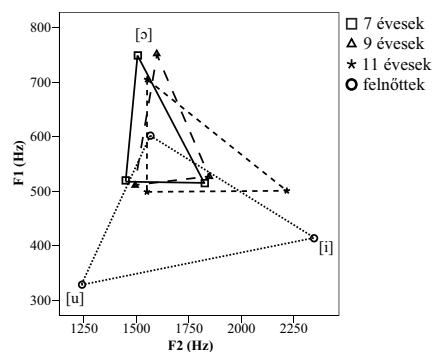
Következtetések

A jelen kutatásban a 9 és 11 éves gyermekek magánhangzóinak első és második formánsainak frekvenciaértékeit, valamint a magánhangzók időtartamát elemeztük. Vizsgálatunk célja az volt, hogy a magánhangzók akusztikai-fonetikai szerkezetében megfigyelhető változásokat feltérképezzük az életkor előrehaladtával. Választ kerestünk arra a kérdésre, vajon melyik életkori szakaszban válnak a gyermekek magánhangzói hasonlónak a felnőttekéihez.

Hipotéziseink részben igazolódtak. Szignifikáns különbség van az általunk vizsgált 9 és 11 éves gyermekek magánhangzói között mind a formánsok frekvenciaszerkezetében, mind az időtartamban. Az adatok azt mutatják, hogy a gyermekek magánhangzóinak első két formánsa 11 éves kor után válik csak hasonlónak a felnőttekéhez. Hasonló eredményt kaptunk a magánhangzók időtartamának tekintetében is. Feltételezhetően a lassabb artikulációs tempó miatt a magánhangzók 9 éves korban még hosszabb időtartamban realizálódnak, mint a 11 éveseknél. Az életkor előrehaladtával ugyan a gyermekek egyre gyakorlottabb beszélőkké válnak, fokozatosan csökken a gyermekek és a felnőttek magánhangzóinak akusztikai-fonetikai szerkezete közötti különbség, de az adatok azt jelzik, hogy még az anyanyelv-elsajátítás későbbi időszakában sem fejeződik be a magánhangzók stabilizálódásának folyamata.

Minden életkorban adatoltunk bizonyos, de nem jelentős mértékű (a legtöbb esetben nem szignifikáns) különbségeket a lányok és a fiúk magánhangzóinak F_1 -, F_2 - és időtartamértékeiben. Ezek a fiúk és lányok közötti biológiai, fiziológiai, szociofonetikai, kulturális tényezők különbségével magyarázhatók.

Kutatásunk eredményeit összevetve korábbi, 7 évesekkel (Auszmann 2014) és felnőttekkel (Grácsi–Horváth 2010) végzett kutatások eredményeivel (lásd 6. ábra) azt mondhatjuk, hogy a 11 éves gyermekek adatai már nem a fiatalabb gyermekek frekvenciaértékeivel mutatnak hasonlóságot, ugyanakkor még nem azonosak a felnőttek spontán beszédében mérhetőekkel.



6. ábra

A magánhangzótér változása az életkor előrehaladtával

Megfigyelhető az, hogy megindult egy változási folyamat a magánhangzók akusztikai szerkezetében a magyarra jellemző nyelvspecifikus értékek felé.

A gyermeknyelvi magánhangzók vizsgálata azért fontos, mert az akusztikai stabilizálódásuknak a folyamatának feltárásán keresztül az anyanyelv-sajátítás menetét is jobban megismerjük. Képet kapunk arról is, hogy milyen akusztikai-fonetikai jellemzői vannak a beszédnek tipikus fejlődés esetén. Ezek az eredmények továbbá fontos összehasonlítási alapul szolgálhatnak az atipikus beszédfejlődésű gyermekek vizsgálatához.

Irodalom

- Auszmann Anita 2014. Magánhangzók akusztikai szerkezete 7 és 9 éves kisiskolások spontán beszédében. In Bátyi Szilvia – Navracsics Judit – Vigh-Szabó Melinda (szerk.): *Nyelvelsajátítási-, nyelvtanulási- és beszéd kutatások. Pszicholingvisztikai tanulmányok IV.* Gondolat Kiadó, Budapest–Veszprém. 90–97.
- Auszmann, Anita – Neuberger, Tilda 2014. Age- and gender-related differences in formant structure during the stabilization process of vowels. In: *Proceedings of the Olomouc Linguistics Colloquium 2014.* 663–676.
- Bata Sarolta 2007. *Időszerkezeti mintázatok a spontán beszédben.* Szakdolgozat. ELTE, Budapest.
- Boersma, Paul – Weenink, David 2011. Praat: Doing phonetics by computer (Version 5.3). Elérhető: http://www.fon.hum.uva.nl/praat/download_win.html (A letöltés ideje: 2011. október 10.)
- Bóna Judit 2009. Az idős életkor tükröződése a magánhangzók ejtésében. *Beszéd kutatás 2009.* 76–87.
- Bóna Judit – Imre Angéla 2010. A rövid-hosszú magánhangzók óvodás és kisiskolás gyermekek beszédprodukcójában. In Navracsics Judit (szerk.): *Nyelv, beszéd, írás. Pszicholingvisztikai tanulmányok I.* Segédkönyvek a nyelvészet tanulmányozásához 107. Tinta Könyvkiadó, Budapest. 49–56.
- Deme Andrea 2012. Óvodások magánhangzóinak akusztikai jellemzői. In Markó Alexandra (szerk.): *Beszédtudomány: Az anyanyelvelsajátítástól a zöngékezdési időig.* ELTE Bölcsészettudományi Kar – MTA Nyelvtudományi Intézet, Budapest. 77–99.
- Fant, Gunnar. 1966. *A note on vocal tract size factors and non-uniform F-pattern scalings.* Speech Transmission Laboratory Quarterly Progress and Status Report 1. 22–30.
- Gósy Mária 2003. A spontán beszédben előforduló megakadásjelenségek gyakorisága és összefüggései. *Magyar Nyelvőr* 127. 257–277.
- Gósy Mária 2004. *Fonetika, a beszéd tudománya.* Osiris Kiadó, Budapest.
- Gósy Mária 2005. *Pszicholingvisztika.* Osiris Kiadó, Budapest.
- Gósy Mária 2006. A beszédhangok megkülönböztetésének fejlődése. *Beszéd kutatás 2006.* 147–159.
- Gósy Mária – Beke András 2010. Magánhangzó-időtartamok a spontán beszédben. *Magyar Nyelvőr* 134. 140–165.
- Gráczi Tekla Etelka – Horváth Viktória 2010. A magánhangzók realizációja spontán beszédben. *Beszéd kutatás 2010.* 5–16.

- Gyarmathy Dorottya 2007. Az alkohol hatása a beszédprodukción. *Beszéd kutatás* 2007. 108–120.
- Horváth Viktória 2013. Temporális szerveződés kilencéves gyermekek spontán beszédében. *Beszéd kutatás* 2010. 144–159.
- Huber, Jessica L. – Stathopoulos, Elaine T. – Curione, Gina M. – Ash, Theresa A. – Kenneth, Johnson 1999. Formants of children, women, and men: the effects of vocal intensity variation. *Journal of the Acoustical Society of America* 106/3. 1532–1542.
- International Phonetic Association 1999. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University, Cambridge.
- Laczkó Mária 2009. Középiskolai tanulók beszédének temporális jellemzői. *Magyar Nyelvőr* 133. 447–467.
- Laczkó Mária 2011. Óvodások és kisiskolások spontán mondatalkotási folyamatai. *Magyar Nyelvőr* 135. 440–458.
- Lee, Sungbok – Potamianos, Alexandros – Narayanan, Shrikanth 1999. Acoustics of children's speech: developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America* 105/3. 1455–1468.
- Neuberger Tilda 2011. Gyermekek spontán beszédének szerkesztettsége és folyamatossága. *Beszéd kutatás* 2011. 83–95.
- Neuberger Tilda 2014. *A spontán beszéd sajátosságai gyermekkorban*. ELTE Eötvös Kiadó, Budapest.
- Olaszy Gábor 2006. *Hangidőtartamok és időszerkezeti elemek a magyar beszédben*. Nyelvtudományi Értekezések 155. Akadémiai Kiadó, Budapest.
- Perry, Theodore L. – Ohde, Ralph N. – Ashmead, Daniel H. 2001. The acoustic bases for gender identification from children's voices. *Journal of the Acoustical Society of America* 109/6. 2988–2998.
- Vorperian, Houry K. – Kent, Ray D. 2007. Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research* 50/6. 1510–1545.
- Whiteside, Sandra P. – Hodgson, Carolyn 2000. Some acoustic characteristics in the voices of 6-to 10-year-old children and adults: a comparative sex and developmental perspective. *Logopedics Phoniatrics Vocology* 25/3. 122–132.
- Whiteside, Sandra P. – Dobbin, Rachel – Henry, Luisa 2003. Patterns of variability in voice onset time: a developmental study of motor speech skills in humans. *Neuroscience Letters* 374/1. 29–32.

ALKALMAZÁSOK, ESZKÖZÖK

ULTRAHANGOS NYELVKONTÚRKÖVETÉS AUTOMATIKUSAN: A MÉLY NEURONHÁLÓKON ALAPULÓ AUTOTRACE ELJÁRÁS VIZSGÁLATA

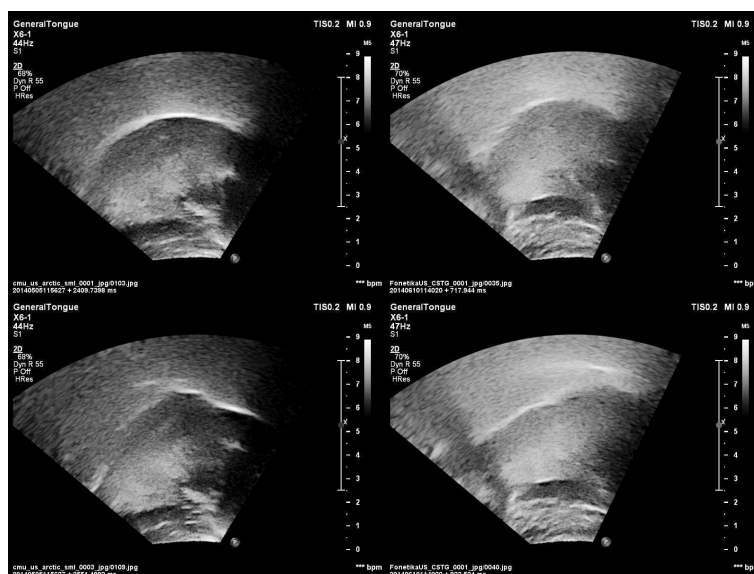
Csapó Tamás Gábor – Csopor Dávid

Bevezetés

Az artikuláció és az akusztikum kapcsolata régóta foglalkoztatja a beszéd-kutatókat. Beszéd közben a nyelv mozgását különböző technológiák segítségével lehet rögzíteni, pl. röntgen (Öhman–Stevens 1963; Bolla 1995), ultrahang (Stone et al. 1983; Stone 2005), EMA – elektromágneses artikulográf (Schönle et al. 1987; Mády 2008), MRI – mágnesrezonancia-képalkotás (Baer et al. 1991; Woo et al. 2012). Az ultrahangos technológia előnye, hogy egyszerűen használható, elérhető árú, valamint nagy felbontású (akár 800×600 pixeles) és nagy sebességű (akár 100 képkocka/s-os) felvétel készíthető vele. Az ultrahang-, MRI- és röntgentechnológiák hátránya viszont ebben a témakörben, hogy a rögzített képsorozatból ki kell nyerni a nyelv körvonalát ahhoz, hogy az adatokon további vizsgálatokat lehessen végezni.

Az 1. ábra néhány példát mutat a nyelvről rögzített ultrahangfelvétellel. A felvételeken bal oldalon látható a nyelvgyök, jobb oldalon a nyelvhegy; a kettő között a nyelv felső felülete. A felvételek során az ultrahangtranszducert az áll alá helyezik; így az ultrahangjelben a legnagyobb változást a nyelv izomzatának felső határa okozza, ami az ultrahangos képeken ideális esetben jól kivehető fehér sávot eredményez. Mivel a hullámok nagy része nem jut tovább a nyelv felső határán, így a távolabbi szövetpontokról, a szájpadról kevesebb az információnk. Az 1. ábrán látható, hogy a képek minősége széles skálán mozog, mivel az ultrahang-technológia nem mindig nyújt teljesen tökéletes nyelvkontúrt. A bal felső képen jól kivehető a nyelv kontúrja; ezzel szemben a bal alsó képen a kontúr nem folytonos, hanem szakadás vagy ugrás látható. A jobb felső képen a nyelvkontúr kevésbé erőteljesen látszik; a jobb alsó képen pedig több kontúr is látható egymás felett, amit valószínűleg a szájpadról közelsége okoz. A kép minősége függ a beszélőtől, általában fiatalabbnál és nőknél jobb, de ez függ a száj hidratációjától is (Stone 2005).

Ennek oka egyrészt az lehet, hogy a nőknek kisebb az állkapcsa, ezért a szájüreg nagyobb részére terjed ki a felvétel; másrészt idősebbeknél nagyobb mértékű az áll elzsírosodása, ami nehezíti a nyelvkontúr követését.



1. ábra

Ultrahangfelvételek a nyelvről (A képeken bal oldalon látható a nyelvgyök, jobb oldalon a nyelvhegy; a kettő között a nyelv felső felülete.)

Az ultrahangos nyelvkontúrkövetés hagyományosan manuális vagy félautomatikus módon történt, azonban az elmúlt időszakban automatikus megoldások is megjelentek erre a célra.

Az **EdgeTrak** volt az első nyelvkontúrkövető program, amelyet széles körben kezdtek el használni (Li et al. 2005). A módszer a klasszikus „kígyókövető” algoritmusokon alapul (Akgul et al. 1999), és jelenleg az automatikus ultrahangos nyelvkontúr *de facto* szoftverének számít. A **Palatoglossatron** alapvető célja, hogy a szájpadróláshoz képest legyen lehetőség megállapítani a nyelv helyzetét (Mielke et al. 2005). A programban alkalmazott GLoSsatron félautomatikus nyelvkontúrkövetés a képben lévő gradiensek (hirtelen fényerőváltozás) mérésén alapul (Baker et al. 2005). Az **Ultra-CATS** program a manuális kontúrkövetés sebességének növelésére készült, de rendelkezik félautomatikus nyelvkontúrkövető funkcióval is (Bressmann et al. 2005). A **TongueTrack** módszer a fentiekkel szemben az egymás után következő ultrahangképek térbeli és időbeli kapcsolatát próbálja meg kihasználni, hiszen a nyelv mozgása korlátozott sebességű (Tang et al. 2012).

Az **AutoTrace** szoftver az automatikus nyelvkontúrkövetésre mély neuronhálókat, ezen belül translational Deep Belief Network-öt (tDBN) alkalmaz (Fasel–Berry 2010). Mivel ez géptanulás-alapú eljárás, ezért először jelentős méretű tanítóadatra, azaz manuálisan berajzolt nyelvkontúrra van szükség. Berry és munkatársai (2012) javasolnak egy módszert, amellyel nagy mennyiségű képből kiválasztható az az optimális részhalmaz, amelyet a gépi tanuláshoz érdemes felhasználni. Sung és munkatársai (2013) azt találták, hogy a nyelvkontúrkövetés pontossága növelhető az „iteratív újratanítás” eljárással is. Az AutoTrace a <https://github.com/jjberry/Autotrace> honlapról szabadon letölthető (Hahn–Powell–Archangeli 2014). A legújabb kutatási eredmények szerint a manuális nyelvkontúrkövetés átlagos négyzetes hibája 7 pixel, és a fenti automatikus algoritmusok közül az EdgeTrak, a TongueTrack és az AutoTrace eljárások nyújtják a legpontosabb nyelvkontúrkövetést (Csapó–Lulich 2014a, 2014b).

A tanulmányban bemutatjuk legújabb eredményeinket, melyekhez az AutoTrace eljárás automatikus nyelvkontúrkövetését vizsgáltuk (Csopor 2014). A manuális és az automatikus kontúrok hibájának mérésére hagyományosan alkalmazott átlagos négyzetes hiba mellett új típusú hibamértékeket vezetünk be, melyek jobban mérik az automatikusból hiányzó vagy esetlegesen hozzáadott nyelvkontúrszakaszokat. Emellett meghatározzuk, hogy a tanítóadat mennyiségének függvényében milyen mértékben tudja az automatikus nyelvkontúrkövetés a manuálist közelíteni.

Anyag, módszer

A továbbiakban bemutatjuk a kutatás során rögzített új ultrahangos beszédadatbázist; valamint a manuális és automatikus nyelvkontúrkövetés körülményeit, továbbá ezek összehasonlításának módszerét.

Felvételi körülmények

Egy amerikai angol (jelölés: EN1) és egy magyar anyanyelvű (jelölés: HU1) férfi beszélőtől rögzítettünk felvételeket az Indiana University Speech Production Laboratory csendesített szobájában (Csapó–Lulich 2014b, 2014c). EN1 beszélő a CMU-ARCTIC adatbázis első 135 mondatát (Kominék–Black 2003), míg HU1 beszélő a PPBA adatbázis első 210 mondatát olvasta fel (Olaszy 2013). A beszédfelvételek Shure kondenzátormikrofonnal készültek, melyet 48 kHz-en egy National Instruments jelfeldolgozó kártyával digitalizáltunk. A nyelv mozgását Philips EpiQ-7G ultrahangos rendszerrel és xMatrix 6-1 MHz transzdúccerrel rögzítettük. A felvételek során az ultrahangfej elmozdulásának elkerülése végett egy speciálisan erre a feladatra kialakított sisakot alkalmaztunk (típusa: Ultrasound Stabilisation Headset, Articulate Instruments Ltd), mellyel az ultrahangtranszdúcer szorosan a beszélő álla alá rögzíthető (Wrench 2007). Az ultrahanggépet a háttérzaj csökkentésének céljából a süketszobán kívülre helyeztük, és csak az ultrahangfejet és -kábelt vezettük a beszélőhöz. Az arc alsó részéről (így az ajkak mozgásáról) videofel-

vétel készült, valamint a hangszalagok rezgéséről EGG-felvételt rögzítettünk; ezeket azonban a jelen kutatásban nem használtuk fel.

Az ultrahangadatok eredetileg DICOM formátumban készültek 800×600 pixel felbontásban és 40–50 kép/s közötti sebességgel. A DICOM fájlokat az Image-J programmal (National Institutes of Health, USA, v1.46a, <http://imagej.nih.gov/ij>) JPG képekké alakítottuk.

Az elkészült párhuzamos beszéd – nyelvultrahang – videó – EGG adatbázis kutatási célra szabadon elérhető és felhasználható (Csapó–Lulich 2014c).

Manuális nyelvkontúrkövetés

A manuális nyelvkontúrkövetést a fenti felvételek egy kisebb részén (EN1 beszélő: 8 mondat, HU1 beszélő: 9 mondat) egy hallgató végezte egy speciálisan erre a feladatra készült weboldal segítségével. A weboldal az Indiana University, Speech Production Laboratoryban készült, és a lokális hálózaton elérhető. A hallgató a feladat elvégzése előtt általános ismertetést kapott a képek orientációjáról (pl. nyelvhegy a jobb oldalon; a nyelv hátulja a bal oldalon; a jobb oldali sötét folt az állkapocs csontja), illetve utasításokkal láttuk el, hogy az 1. ábrához hasonló helyzetekben hogyan végezze a nyelvkontúrok berajzolását. Minden esetben a nyelv felső részére utaló fehér sáv alján történt a kontúr berajzolása. A feladat elvégzése összesen 5-6 órányi időt vett igénybe. A nyelvkontúrt képenként kb. 200 ponttal reprezentáltuk. A manuális nyelvkontúrkövetés eredménye SQL adatbázisba került, melyből CSV formátumba exportálva lehet az adatokat feldolgozni.

Az EN1 beszélőtől összesen 1140 ultrahangos képet, míg a HU1 beszélőtől összesen 1457 képet használtunk fel a kísérletekben.

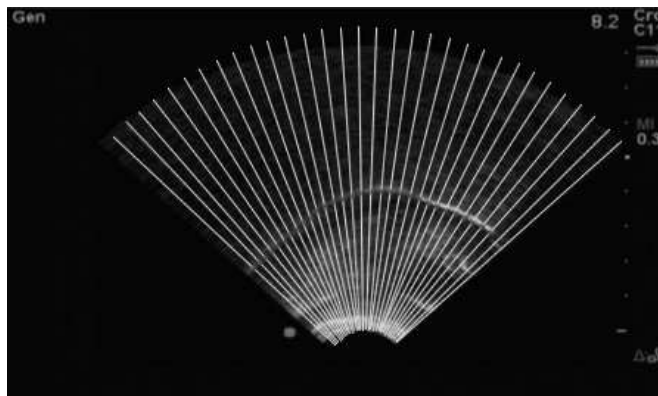
Automatikus nyelvkontúrkövetés

Az automatikus nyelvkontúrkövetést az AutoTrace programmal végeztük. A 800×600 pixel méretű képeket először átméreteztük 550×413 -re, majd eltoltuk vízszintesen 140 pixellel, függőlegesen 30 pixellel, végül 720×480 pixel méretű képet kapva (az origó a bal felső sarokban volt). Az átméretezés és eltolás célja, hogy az AutoTrace-ben definiált polárkoordináta-rendszer illeszkedjen a képeken az ultrahang által rögzített területhez (2. ábra). A képekből kiválasztottuk azt a részt, amely várhatóan tartalmazza a nyelv teljes szakaszát. A manuális nyelvkontúr pontsorozatát interpoláltuk az AutoTrace 32 radiális egyenesére. Az interpoláció során az adott poláregyeneshez tartozó szomszédos pontokat súlyoztuk aszerint, hogy a polárkoordinátára konvertált pontok szöge mennyire van közel az adott egyeneshez. Ennek célja a kvantálási hiba csökkentése volt (vö. Csopor 2014).

Az AutoTrace-ben az automatikus nyelvkontúrkövetéshez egy betanító lépésre van szükség. Tanítóadatként először EN1 beszélő első mondatát használtuk fel (összesen 169 kép), majd az eredményként kapott betanított neurális háló modellel a 8. mondaton végeztünk tesztelést. A tanítóadat méretét fokozatosan, mondatonként növeltük; majd a tesztelést ugyanazon a mondaton végeztük. Ezután HU1 beszélő ultrahangos képeivel ismételtük meg a fenti

kísérletet. Minden tanítási és tesztelési lépés után megvizsgáltuk az automatikus nyelvkontúrkövetés pontosságát, melyet a következő fejezetben ismertett hibamértékekkel számítottunk.

Az AutoTrace mélyneuronháló-alapú gépi tanulást végez, melyre a translational Deep Belief Network architektúrát alkalmazza. A neurális hálózatban alapértelmezésben két rejtett réteget alkalmaznak (Fasel–Berry 2010).



2. ábra

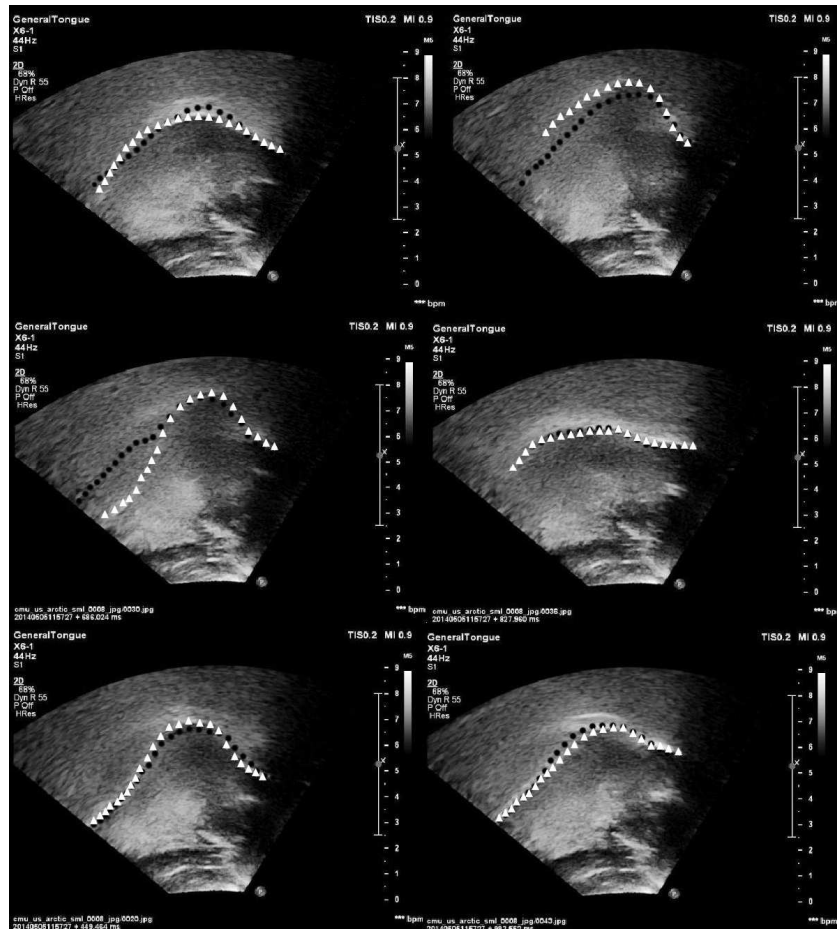
Az AutoTrace-ben definiált polárkoordináta-rendszer (A nyelvkontúr és a 32 radiális egyenes metszéspontjai az AutoTrace-ben eltárolt értékek. Forrás: <https://github.com/jjberry/Autotrace>.)

Az összehasonlítás módszere

Az automatikus nyelvkontúrkövetés tipikus hibái a 3. ábrán láthatóak: fehér háromszögekkel jelöltük a manuálisan berajzolt nyelvkontúrt; míg fekete körökkel az AutoTrace-ből származó automatikus kontúrt. Referenciának a manuális kontúrt használjuk. Három fő hibatípust állapítottunk meg: a) az automatikus kontúr bizonyos távolságnál messzebb van a referenciától (pl. bal középső kép bal oldala a 3. ábrán); b) a referenciában egy adott szakaszon nincs berajzolt pont, de az automatikus kontúrban van (pl. jobb felső kép bal oldala a 3. ábrán); c) a referenciában egy adott szakaszon van berajzolt pont, de az automatikus kontúrban nincs (pl. jobb középső kép bal oldala a 3. ábrán). A nyelvkontúrkövetés hibáinak mérésére négy hibamértéket alkalmazunk és vizsgálunk: (i) átlagos négyzetes hiba (Root Mean Squared Error, RMSE), (ii) két kontúr közötti terület (Area), (iii) hozzáadások aránya (Insertion), (iv) törlések aránya (Deletion).

A szakirodalomban általában az RMSE hibamértéket vagy ehhez hasonlót (például átlagos abszolút hiba) alkalmaznak (Li et al. 2005; Sung et al. 2013; Hahn-Powell–Archangeli 2014). Az RMSE számítása során páronként vesz-

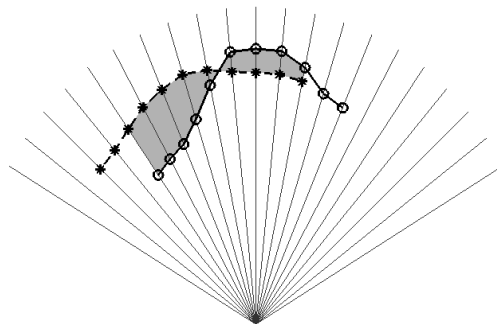
szük a manuális és automatikus kontúrban lévő pontokat; koordinátaikat egymásból kivonjuk; ezen távolságokat négyzetre emeljük, majd összegezzük; végül az összegből gyököt vonunk. Azokat a pontokat nem vesszük figyelembe az RMSE számításánál, ahol nincs automatikusan és manuálisan is meghatározott pont. Az RMSE tehát jól méri az a) típusú hibát, de ugyanakkor nem jellemzi a b) és c) típusú hibákat.



3. ábra

Néhány példa manuálisan berajzolt (fehér háromszögek) és automatikusan előállított (fekete körök) nyelvkontúrra

Az Area hibamérték számítását a 4. ábra mutatja (a szürkével jelölt terület). A második paraméterünk bevezetésének motivációja az volt, hogy az RMSE-vel szemben ez várhatóan pontosabban méri a manuális és automatikus nyelvkontúr közötti hibát (Csopor 2014). A pontokat a számítás során egyenes szakaszokkal kötöttük össze, így meghatározva a két kontúr közötti területet. Az RMSE-hez hasonlóan itt is csak akkor lehet területet számolni, ha adott poláregyenesen van automatikusan meghatározott és manuálisan be rajzolt pont is, azaz az Area szintén az a) típusú hibát méri.



4. ábra

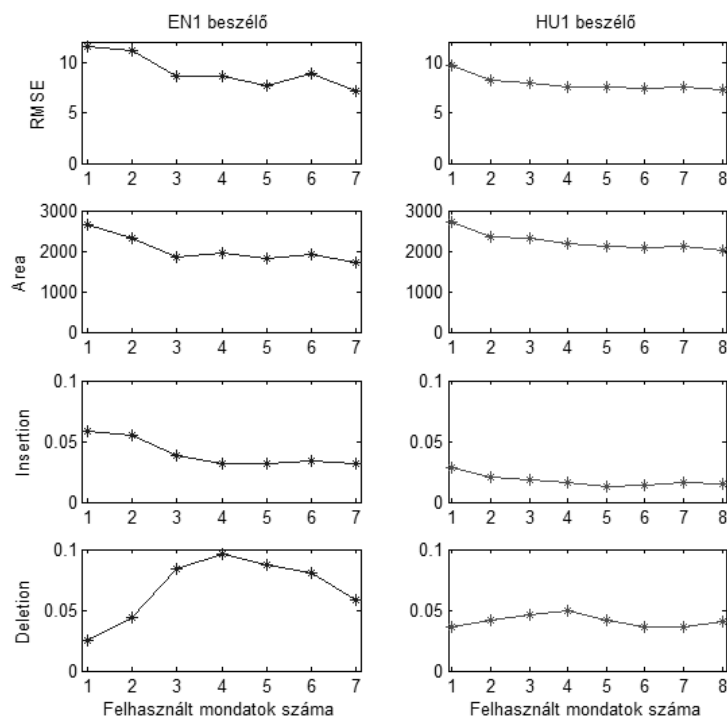
Az Area hibamérték számítása két nyelvkontúr között
(Manuális: folytonos vonal, körök; automatikus: szaggatott vonal, csillagok.)

A b) és c) típusú hibák méréséhez a beszédfelismerésben használt Word Error Rate-hez hasonlóan kialakítottuk a Tracking Error Rate (TER) változóhármast, mely tartalmazza az Insertiont (beillesztést), a Deletion (törlést) és a Substitutiont (helyettesítést) (Csopor 2014). A Substitution az RMSE-nek megfelelő, ezért azzal a továbbiakban nem foglalkozunk. A TER-Insertion számítása során megszámoljuk azokat a pontokat, melyek az automatikus kontúrban megjelennek, de a referenciában nem; majd ezt elosztjuk az összes referenciában lévő pontok számával. A TER-Deletion számítása épp fordított: megszámoljuk azokat a pontokat, melyek az automatikus kontúrban nem jelennek meg, de a referenciában igen; majd ezt elosztjuk az összes referenciában lévő pontok számával. Mivel a TER-Insertion és a TER-Deletion is arányszámok, ezért értéküket százalékban adjuk meg.

Eredmények

A mérési eredmények az 5. ábrán láthatóak, amelyen a négyféle hiba változását ábrázoltuk a tanítóadat mennyiségének függvényében, külön a két beszélőre. Megfigyelhető, hogy az RMSE és az Area mindkét beszélő harmadik modelljénél (azaz amikor összesen 3 mondatot, kb. 600 képet használunk tanításra) beáll egy bizonyos értékre, és egyre alacsonyabb mértékben csökken

további tanítóadat hozzáadásával. Mind EN1, mind HU1 beszélő esetén három mondatot felhasználva az RMSE 8 pixelre csökken, amely közel meg-
 egyezik a manuális nyelvkontúrkövetés átlagosan 7 pixel hibájával (Csapó-
 Lulich 2014b). Az Area hibamérték az RMSE-hez hasonló tendenciát mutat:
 a két hibamérték között a korreláció magas (EN1: 0,97; HU1: 0,99).



5. ábra

A négyféle hibamérték változása a tanítóadat mennyiségének függvényében a két beszélőre

Az Insertion és Deletion értékei nem változnak egyértelműen a tanítóadat növekedésével. Az Insertion először csökken az ötödik mondat hozzáadásáig, majd kis mértékben növekszik. A Deletion hiba a negyedik mondat hozzáadásáig növekszik mindkét beszélő esetén (EN1-nél egészen 9,5%-ig); további tanítóadat hozzáadásával pedig csökkenni kezd. Ezen két hibamérték értékei valószínűleg nem közvetlenül függnek a tanítóadat mennyiségétől. Az Insertion és Deletion típusú hibák csökkentéséhez a tanítóadat előzetes változtatása vagy más neurális háló architektúra lehet szükséges.

Következtetések

Az eredmények szerint az AutoTrace programban két rejtett rétegből álló translational Deep Belief Network típusú mély neuronháló architektúrával, kb. 600 ultrahangos képet tanítóadatként használva már közelíthető a manuális nyelvkontúrkövetés átlagos négyzetes hibaként mért pontossága. Az átlagos négyzetes hiba a tanítóadat mennyiségét növelve folyamatosan csökken. Az automatikusan meghatározott nyelvkontúrszakasz hossza is változik a tanítóadat függvényében: eredményeink szerint a tanítóadat felét felhasználva volt a legrövidebb a nyelvkontúr. Előzetes kísérleteink alapján az Insertion és Deletion típusú hiba a neurális hálózat architektúrájának változtatásával (pl. rejtett rétegek számának növelése) tovább csökkenthető (Csopor 2014).

Az automatikus nyelvkontúrkövetés a beszéd kutatás alapkérdéseinek megválaszolása mellett hasznos lehet a nyelvoktatásban, a beszédrehabilitációban, illetve a beszédtechnológiában, audiovizuális beszéd szintézisben is (Hueber et al. 2011). A további kutatás során tervezzük a fenti adatbázis képeiből ki nyerhető nyelvkontúr alakját összehasonlítani a beszédjelből mérhető formánsértékekkel. A fenti kutatásban a kétdimenziós ultrahang-technológiát alkalmaztuk, melynek során a nyelv középvonaláról rendelkezünk információval. A legújabb, háromdimenziós technológiával azonban a nyelv teljes felületéről információt kaphatunk beszéd közben. Előzetes kísérleteink szerint a nyelv mozgása közel sem szimmetrikus, így a háromdimenziós vizsgálattal olyan kérdésekre is válasz kapható, hogy a nyelv mozgása és alakja milyen mértékben járul hozzá az akusztikai kimenet formálásához.

Az elkészült párhuzamos beszéd – nyelvultrahang – videó – EGG adatbázis kutatási célra szabadon elérhető és felhasználható (Csapó–Lulich 2014c).

Irodalom

- Akgul, Yusuf Sinan – Kambhamettu, Chandra – Stone, Maureen 1999. Automatic extraction and tracking of the tongue contours. *IEEE Transactions on Medical Imaging* 18/10. 1035–1045.
- Baer, Thomas – Gore, John C. – Gracco, L. C. – Nye, Patrick W. 1991. Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels. *The Journal of the Acoustical Society of America* 90/2. 799–828.
- Baker, Adam – Mielke, Jeff – Archangeli, Diana 2005. *Tracing the tongue with GLoSsatron*. Előadás. Ultrafest III. Tucson, AZ, USA. <http://apil.arizona.edu/presentations/GLoSsatronPresentation.ppt>. (A letöltés ideje: 2015. február 1.)
- Berry, Jeff – Fasel, Ian – Fadiga, Luciano – Archangeli, Diana 2012. Training deep nets with imbalanced and unlabeled data. In: *Proceedings of Interspeech*. Portland, OR, USA. 1756–1759.
- Bolla, Kálmán 1995. *Magyar fonetikai atlasz. A szegmentális hangszerkezet elemei*. Nemzeti Tankönyvkiadó, Budapest.
- Bressmann, Tim – Heng, Chiang-Le – Irish, Jonathan C. 2005. Applications of 2D and 3D ultrasound imaging in speech-language pathology. *Journal of Speech-Language Pathology and Audiology* 29/4. 158–168.

- Csapó, Tamás Gábor – Lulich, Steven M. 2014a. *Comparison of tongue contour extraction methods from ultrasound images for use in text-to-speech synthesis*. Előadás. Inaugural Conference of the Hungarian Cultural Association. Bloomington, IN, USA. <http://speechlab.tmit.bme.hu/csapo/downloads/CsapoLulich-HungStudies2014-presentation.pdf>. (A letöltés ideje: 2015. február 1.)
- Csapó, Tamás Gábor – Lulich, Steven M. 2014b. Tongue contour tracings from 2D ultrasound image sequences: quantification of measurement error using manual and automatic tracing methods. (Előkészületben.)
- Csapó, Tamás Gábor – Lulich, Steven M. 2014c. IU_ULTRASOUND data set. http://speechlab.tmit.bme.hu/csapo/IU_ULTRASOUND/. (A letöltés ideje: 2015. január 7.)
- Csopor Dávid 2014. Mély neuronhálók alkalmazása ultrahangos nyelvkontúr követésre. TDK-dolgozat. BME VIK. <https://tdk.bme.hu/VIK/DownloadPaper/Mely-neuronhalok-alkalmazasa-ultrahangos>. (A letöltés ideje: 2014. november 5.)
- Fasel, Ian – Berry, Jeff 2010. Deep Belief Networks for Real-Time Extraction of Tongue Contours from Ultrasound During Speech. In: *Proceedings of ICPR*. Istanbul, Turkey, 1493–1496.
- Hahn-Powell, Gustave V. – Archangeli, Diana 2014. AutoTrace: An automatic system for tracing tongue contours. *The Journal of the Acoustical Society of America* 136/4. 2104.
- Hueber, Thomas – Benaroya, Elie-Laurent – Denby, Bruce – Chollet, Gérard 2011. Statistical mapping between articulatory and acoustic data for an ultrasound-based silent speech interface. In: *Proceedings of Interspeech*. Florence, Italy. 593–596.
- Kominek, John – Black, Alan W. 2003. *CMU ARCTIC databases for speech synthesis*. Carnegie Mellon University.
- Li, Min – Kambhamettu, Chandra – Stone, Maureen 2005. Automatic contour tracking in ultrasound images. *Clinical Linguistics and Phonetics* 19/6-7. 545–554.
- Mády Katalin 2008. Magyar magánhangzók vizsgálata elektromágneses artikulográfiával normál és gyors beszédben. *Beszédkutató* 2008. 52–66.
- Mielke, Jeff – Baker, Adam – Archangeli, Diana – Racy, Sumayya 2005. Palatron: A technique for aligning ultrasound images of the tongue and palate. *Coyote Papers* 14. 97–108.
- Olaszy Gábor 2013. Precíziós, párhuzamos magyar beszédadatbázis fejlesztése és szolgáltatásai. *Beszédkutató* 2013. 261–270.
- Öhman, Sven – Stevens, Kenneth N. 1963. Cineradiographic studies of speech: Procedures and objectives. *The Journal of the Acoustical Society of America* 35. 1889.
- Schönle, Paul W. – Gräbe, Klaus – Wenig, Peter – Höhne, Jörg – Schrader, Jörg – Conrad, Bastian 1987. Electromagnetic articulography: Use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract. *Brain and Language* 31/1. 26–35.
- Stone, Maureen – Sonies, Barbara C. – Shawker, Thomas H. – Weiss, George – Nadel, Lawrence 1983. Analysis of real-time ultrasound images of tongue configuration using a grid-digitizing system. *Journal of Phonetics* 11. 207–218.
- Stone, Maureen 2005. A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics and Phonetics* 19/6-7. 455–501.

- Sung, Jae-Hyun – Berry, Jeff – Cooper, Marissa – Hahn-Powell, Gustave – Archangel, Diana 2013. Testing AutoTrace: A machine-learning approach to automated tongue contour data extraction. *Ultrafest VI*. Edinburgh, UK. 9–10. http://www.qmu.ac.uk/casl/conf/ultrafest_2013/docs/G_HahnPowel_1_ultrafest.pdf. (A letöltés ideje: 2015. február 1.)
- Tang, Lisa – Bressmann, Tim – Hamarneh, Ghassan 2012. Tongue contour tracking in dynamic ultrasound via higher-order MRFs and efficient fusion moves. *Medical Image Analysis* 16/8. 1503–1520.
- Woo, Jonghye – Murano, Emi Z. – Stone, Maureen – Prince, Jerry L. 2012. Reconstruction of high-resolution tongue volumes from MRI. *IEEE Transactions on Biomedical Engineering* 59/12. 3511–3524.
- Wrench, Alan 2007. Articulate Assistant Advanced: Ultrasound module. *Ultrafest IV*. New York, NY, USA.

Az első szerzőt részben a Fulbright ösztöndíj támogatta a kutatás során. A szerzők köszönetüket fejezik ki Steven M. Lulichnak (Indiana University, Bloomington, IN, USA), hogy a felvételek a Speech Production Laboratory eszközei segítségével elkészülhettek. A szerzők emellett köszönetet mondanak Elizabeth Mazzoconak (Indiana University, Bloomington, IN, USA) a manuális nyelvkontúrkövetésért, valamint Gustave Hahn-Powellnek (Arizona Phonological Imaging Lab, University of Arizona, Tucson, AZ, USA) az AutoTrace eljárás használatában nyújtott segítségével.

VIZUÁLIS PERCEPCIÓ ÉS NYELVI FELDOLGOZÁS

Hunyadi László – Szekrényes István – Sziklai István

Bevezetés

A jelen tanulmányban olyan kísérletek és próbálkozások eredményeiről, tapasztalatairól számolunk be, amelyek bizonyos lényeges elemeiben akár nehezen értelmezhetőkné is tűnhetnek a fonetika és a fonológia szorosabban vett megközelítésében, azonban amelyek ugyanakkor tanulságokkal is szolgálhatnak a beszédfeldolgozás általános mechanizmusainak jobb megértése szempontjából.

A két éve folyó, több résztvevős, inter- és multidiszciplináris kutatás középpontjában az a problematika áll, milyen módon lehetne elősegíteni a születésüktől fogva siket vagy erősen nagyothalló gyermekek magyar nyelvi beszédfejlődését, különös tekintettel a hangzók helyes kiejtésére és a beszéd prozódijára. Az általános tapasztalat szerint ugyanis a halláson keresztül kapott visszajelzés teljes vagy részleges hiánya jelentősen megnehezíti a természe-

teshez közelítő beszédprodukció kialakítását, aminek a birtokában pedig könnyebbé és akadálymentesebbé válhatna a hallássérültek kommunikációja a halló többséggel, és így sikeresebben, teljesebben tudnának beilleszkedni a környező társadalomba. Szándékunk az, hogy a beszéd fejlesztését összekapcsoljuk a halláspercepció valamiféle elérhetővé tételével, amin keresztül a siketek felhasználhatnák a hallók esetében természetesnek tekinthető akusztikus visszajelzés korrekciós funkcióját. E szándékot követve nem elégedhetünk meg azzal, hogy a napjainkban már egyre hatékonyabb beszéd > szöveg nyelvtchnológiát alkalmazzuk siketek számára olvashatóvá téve a hallott beszédet, hiszen ezzel csak egy egyirányú, a siketek felé mutató beszédkommunikációt segítenénk elő, magát a beszédprodukciót nem. A mindkét irányba ható megoldások között ismert, hogy bizonyos esetekben műtéti beavatkozással (cochleáris implantátummal) is lehet javítani a hallásélményt (vö. Zahnert 2011), azonban a világ siket népességének egy jelentős része számára kulturális, társadalmi, egészségügyi és egyéb okokból ez nem alkalmas; ők a jelnyelvet használják, ami azonban a halló többség számára jobbra nem elérhető (vö. Bartha et al. 2006; Hattyár 2009; Weisleder 2012; Baertschi 2013), és ugyancsak nincs hatással a beszédprodukcióra. Többféle módszer áll rendelkezésre a siketek beszédtanítására akusztikus visszajelzés nélkül is (a magyar nyelv vonatkozásában kiemelendő a Beszédmester és az Aranykapu), azonban a beszéd, ezen belül az egyes beszédhangok specifikus fizikai jellemzőinek a közvetlen, azaz valamilyen módon a halláspercepciót helyettesítő megtapasztalása meglévő agyi funkciók átprogramozásával mégis hiányzik ahhoz, hogy a tanítás eredményességét jelentősen növelni lehessen. Ezen kíván segíteni az itt ismertetendő kutatás azzal, hogy egy eddig még nem próbált, új alapkoncepciót fogalmaz meg. Ennek lényege az ún. audiovizuális transzkódolás, ami a siketek által hallás útján nem érzékelhető beszéd vizuális mintázattá (az általunk bevezetett termonilógiában *videogramm*) való átalakítását foglalja magában. Ez az elgondolás azonban csak fontos elméleti kérdések megfogalmazásával és megválaszolásával ültethető át a gyakorlatba.

Feltesszük, hogy e transzkódolás eredményeként létrejövő vizuális ábrázolás a szenzomotoros integráció elvét követve (vö. Hickok et al. 2011) a hallássérültek esetében is elősegíti a beszédfeldolgozást, azaz a beszéd auditoros és motoros reprezentációi közötti átalakításához (vö. Hickok–Poeppel 2004, 2007) hasonlóan létrejöhet az átalakítás a beszéd vizuális és motoros reprezentációi között is. Legjobb tudomásunk szerint elsőként teszünk kísérletet arra, hogy megválaszoljuk, vajon egy audiovizuális transzkóder által elérhetővé válhatnak-e a beszédhangok a vizuális percepcióban és feldolgozásban. Ezen összefüggésben, bár vannak kutatások, pl. a vizuális téri figyelem és a beszédsegmentálás közötti összefüggésre (vö. Faceotti et al. 2010), újdonságnak számíthat a szándékunk, hogy összevessük a hallás- és a látáspercepciót a motoros beszédfeldolgozásban. Így a kutatás egészének alapkérdése az, vajon létezik-e vizuális motoros beszédfeldolgozás. E kérdés megválaszolása

összetett kutatásokat kíván, amelyekben a vizuális hang- és beszédfeldolgozással összefüggő kísérletek mellett az agyi plaszticitás esetleges megjelenését is vizsgáljuk. E tanulmányban a jelenleg is folyó kutatás egyes eddigi eredményeiről adunk számot, figyelmünket a vizuális hangfeldolgozást elősegítő tanulási folyamat elveire irányítva, szándékaink szerint ezáltal közelebb jutva a vizuális és a halláspercepció közötti általános különbségek megismeréséhez. Dolgozatunk az eddigi tanulságokat összegzi.

Az audiovizuális transzkódolás mint eljárás alapja a következő hipotézis:

1. hipotézis: A hangzó beszéd észlelésében megjelenő fonetikai jegyek olyan, fizikailag mérhető akusztikai jelek nyelvi feldolgozásával jönnek létre, amelyeket vizuális jelekké alakítva e kapott jelsorozat vizuális feldolgozása a szükséges és elégséges mértékben reprezentálja e fonetikai jegyeket, ezeken keresztül magát a hangzó beszédet. A szükséges mérték azt jelenti, hogy az ábrázolások az egyes, az adott (esetünkben a magyar) nyelvben meglévő különböző hangtípusokra specifikusak, az elégséges mérték pedig azt, hogy nem specifikusak az egyes beszélők közötti egyedi hangzókülönbségekre.

Az audiovizuális transzkódolás azonban nem merülhet ki abban, hogy megvalósítja a beszéd vizuális mintázattá való átalakítását, meg kell felelnie a beszédpercepció kívánalmainak is. Így fogalmazzuk meg a következő hipotézist:

2. hipotézis: Az audiovizuálisan transzkódolt mintázatot (a továbbiakban: a videogramot) a kognitív vizuális feldolgozás képes a beszédhangok reprezentációjaként értelmezni, azaz képes a mintázatot kisebb egységekre (hangokra) szegmentálni. Így lehetővé teszi az alacsonyabb szintű nyelvi feldolgozást.

E hipotézis teljesülése az alapfeltétele ugyanis annak, hogy a hallássérült gyermek a tanulás során a vizuális mintázatból megkapja azt a hangképzése milyenségére és minőségére mutató visszacsatolást, amely lehetővé teheti számára a hangképzése finomítását, és így a hallók beszédminőségéhez való közelítését. Ezen hipotézis teljesülése azonban még nem jelenti azt, hogy a transzkódolás alkalmas a hangzó beszéd olyan szintű reprezentálására, ami elvezethet a beszédértéshez is. Ehhez a következő hipotézisnek kellene teljesülnie:

3. hipotézis: Az audiovizuális transzkódolás lehetővé teszi a vizuális alapú nyelvi feldolgozást, beleértve a fonémareprezentációt, a szóazonosítást és a mondatértést.

E hipotézis teljesülése különös jelentőséggel bír a nyelvi feldolgozásban: ehhez a hangok fizikai azonosításán túlmenően a nyelvi rendszer mozgósítására is szükség van, ami során szabályokat követő elvonatkoztatásra és értelmezésre kerül sor. A hangokat jelentésmegkülönböztető fonémáknak feleltetjük meg, ezek sorozatából jön létre a szóértelmezés, majd belőlük szintaktikai szabályokkal mondatokat alkotunk (a fonológiai és általában a nyelvi feldolgozás jelentőségére hallássérülteknél vö. von Muenster–Baker 2014).

A fenti alapkoncepció és az annak alapján megfogalmazott hipotézisek vizsgálatát több tudományterület együttes munkájával végezzük. Vizsgálá-

tunk kiterjed a videogramok percepciók tesztelésére, a kiejtés tanításában való alkalmazására és – a folyamatosan bővülő tapasztalatok alapján – a transzkódolt hang reprezentáció finomítására. Az itt következőkben a tanítást megalapozó azon kísérletekről számolunk be, amelyekben a hangzó beszéd transzkódolással történő reprezentálhatóságát kívánjuk tesztelni, és ezáltal feltárni a vizuális nyelvi feldolgozás általános tulajdonságait.

A transzkódolás elméleti kihívásai: lényegkiemelés a vizuális és nyelvi feldolgozásban

Hajlamosak lehetnénk a transzkódolás eredményeként létrejött ábrázolást, de legalábbis az annak alapjául szolgáló eljárást úgy tekinteni, mint ami valamiféle lényegkiemelésre irányul, hiszen intuíciónk szerint a milliszekundumként változó hanghullámból hangértelmező percepciónk számára sem csakis e legapróbb részletek számítanak; a jel fizikai percepciója kellően robusztus, azaz bizonyos határokon belül megengedi egyes, egyébként prototipikusan mérhető deskriptív paraméterértékek aktuális módosulását vagy egyenesen hiányát (gondoljunk csak a telefonnak, különösen a mobiltelefonnak a megértést leginkább nem zavaró csökkentett átviteli frekvenciatarományára, vagy arra, hogy zajos vagy fiziológiailag módosult környezetben a beszédet közvetítő hanghullám bizonyos összetevői a percepció számára elérhetetlenekké válnak). A hanghullám beszédként, ezen belül minimálisan beszédhangok sorozataként való értelmezése túlmutat e fizikai jelfeldolgozáson: ez a fonetikai és a fonológiai feldolgozás területe, amely a hanghullámot egyrészt hangokra (pontosabban az egyedi ejtési változatokon túlmutató hangtípusokra, prototípusosan szerveződő kategóriákra), másrészt az értelem világának ajtaját megnyitó fonémákra szegmentálja; vö. Näätänen et al. 1997. (Ez utóbbi fontosságára utal, hogy a halláspercepcióban is csak azokat a beszédhangokat azonosítjuk, amelyek létező fonémák reprezentációi.) Ha tehát a transzkódolás célja valamiféle lényegkiemelés, akkor annak elő kell segítenie a nyelvi (minimálisan a fonetikai és fonológiai) feldolgozást. A transzkódolásunkban alkalmazott frekvenciaszűrés (a lejjebb részletezendő Bark-skála alapján) ezen lényegkiemelés elemi, minimális eszköze. A hallás azonban ennél jóval összetettebb. Jellemző rá egyfajta redundancia: egy nem determinisztikusan változó környezetben keletkező fizikai jelsorozatból kell eljutni a hangok és a további, nagyobb beszédelemek értelmezéséhez. Ugyanakkor a transzkódolás túlsó oldala, a látáspercepció, melynek középpontjában az alakfelismerés áll, a jelek redukálására törekszik, azaz minél kevesebb jelből és a lehető legegyszerűbb szervezési elvek alapján (vö. van der Helm 2014) igyekszik egy felismerendő alakot generálni. Így a redundancia szükségessége a halláspercepcióban és elkerülendősége a látáspercepcióban két egymással szemben álló dimenziója e két, a transzkódolásban szükségszerűen találkozó modalitásnak.

Ha sikerül is valamiféle lényegkiemelést alkalmaznunk, felmerülhet a kérdés, vajon megmaradhat-e a transzkódolás „csak” bizonyos fizikai természetű

jeleknek egyik modalitásból (hallás) egy másik modalitásba (látás) való megfeleltetésénél, vagy a transzkódolástól elvárható-e a nyelvi feldolgozás elősegítése is. Ahhoz, hogy elégséges legyen csak a fizikai transzkódolás, annak a feltételnek kellene teljesülnie, hogy a transzkódolt vizuális jelek beindítják az alsóbb szintű nyelvi feldolgozást közvetlenül a látáson keresztül. Ellenkező esetben, bár elméletileg ugyancsak lehetséges beszédhangok (hangtípusok) vagy azok sorozatának a beszédsegmentáláson alapuló észlelése, nem lesz lehetséges a fonémaazonosítás, nem aktiválódik a morfofonológia (azaz a beszéd egyetlen lexikális egységnél kisebb egységekre, mint ragokra, képzőkre való alkategorizálása, továbbá a tövváltozatok észlelése). Bár csak bizonyos megszorításokkal, de a szintaxis mint magasabb szintű nyelvi feldolgozás működhet, hasonlóan ahhoz, ahogyan az izoláló nyelvek esetében történik, hiszen a szintaxis alapvető feladata az alsóbb szintű nyelvi feldolgozás során azonosított szavak hierarchikus szerkezetbe való rendezése.

A beszédfolyamat vizuális úton való megközelítése megkerülhetetlenül felveti a vizuális nyelvi feldolgozás olvasással való kapcsolatának a kérdését. A kétutas kaszkád olvasásmodell (DRC, vö. Coltheart–Rastle 1994; Rastle–Coltheart 1998) megkülönbözteti az olvasás két útját, a fonológiai és a lexikális útjait. Úgy tűnik, a valamilyen szempontból ismeretlen szavakat az előbbi, az ismerteket az utóbbi módon dolgozzuk fel. Mivel egy videogramot globálisan is el lehet sajátítani az alkotó elemei szegmentált ismerete nélkül is, a *videogram* szó felismerése tehát történhet lexikai úton is. E kétféle út megkülönböztetését jól példázza az, hogy a diszlexia (vö. Csépe 2006) egyes típusait is vissza lehet vezetni a fonológiai feldolgozási mód sérülésére (vö. Ramus et al. 2003). Ezzel együtt jelentős különbség van a betűk és a videogramok olvasása között. A betűkön alapuló írásrendszereket az adott kultúrákban megalkotásukkor eredendően egy halláson alapuló értelmezés (nyelvi előfeldolgozás) hozta létre, ami ezen kész mintákon keresztül eleve kínálja a beszéd hallók által hallható vagy siketek által látható objektumainak a diszkrét nyelvi egységekre történő szegmentálását (l. betű-hang megfeleltetés). Ezzel szemben a videogramok megtervezése és megalkotása nem támaszkodik ilyen értelmezésre (előfeldolgozásra), azaz a videogramok „értő olvasása” esetében a vizuális jelek percepció nyelvi feldolgozásának jóval korábbról kellene kezdődnie, magával a vizuális jelfolyam nemvizuálisan, fonetikailag és fonológiaiag értelmezendő szegmentálásával.

A nyelvi feldolgozás lehetőségességének kérdése tehát szorosan összekapcsolódik azzal, vajon milyen kognitív folyamatokat tesz elérhetővé a transzkódolás. Ha a vizuális percepció a hangok mögött elérhetővé teszi az értelmet megalapozó fonológiát és ezen keresztül a morfofonológiát, akkor az annak a jele, hogy az agyi plaszticitás elvét követve a látókéreg – miközben továbbra is betölti szerepét a látáspercepcióban, – átveszi a siketknél elvesztett hallókéreg funkcióját is, és utat nyit a „látva hallás” felé (ennek ellenőrzésére képalkotó vizsgálatokat tervezünk). Ellenkező esetben, azaz alsóbb szintű nyelvi feldol-

gozás híján csupán a vizuális feldolgozás működik: ekkor is van lehetőség hangok azonosítására, azonban ezeket mint vizuális objektumokat (formákat) érzékeljük. Ugyancsak van lehetőség arra, hogy az így transzkódolt hanghulámot vizuális objektumok sorozataként szegmentáljuk, és elméletileg lehetővé válik akár egyes szavak vagy több szóból álló mondatok együttes egészsként való értelmezése is. Ami azonban nem lesz lehetséges, az az, ami a fonológia és a morfofonológia hiányából következik: a szavaknál alacsonyabb szinten vizuális feldolgozásunk nem lesz nyelvi szempontból egyszerre gazdaságos és kreatív. A gazdaságosság és a kreativitás egyaránt a nyelv paradigmákon alapuló modelljéből következik: egy paradigma egymástól fonetikai-fonológiai szempontból akár jelentősen is különböző szóalakokat mint egyazon lexéma csak alakjukban különböző változatait tünteti fel, azaz rendezi azonos csoportba (paradigmába – ettől gazdaságos), egyben lehetővé teszi újabb alakok létrehozását (és egy minta szerinti értelmezését) ugyanezen paradigma alapján (ettől kreatív). A vizuális feldolgozás eredendően ettől különböző csoportosítási elveken alapszik (vö. Gestalt-elmélet: Wertheimer 1924, 1925). Talán különösen a hasonlóság, a folytonosság és a szimmetria tekinthetők a vizuális csoportosítás azon elveinek, amelyek a legjobban rámutatnak arra, milyen feladattal kell megküzdenie a látásnak akkor, amikor a transzkódolás eredményeként a felszínen jelentősen különböző vizuális objektumokat kellene mint összetartozókat (egyazon csoportba, a nyelvre vonatkoztatva: egyazon paradigmába tartozókat) értelmeznie. A transzkódolás gyakorlati alkalmazásainak meghatározásánál érdemes ezen szempontokat megfontolni.

Ami a transzkódolás gyakorlati alkalmazhatóságát illeti, e gondolatmenet szerint érdemes egymástól különválasztva vizsgálnunk azt az esetet, amikor a transzkódolás eredményét, a videogramot mint vizuális objektumot dolgozzuk fel és azt, amikor a videogram alapján nyelvi feldolgozást várunk. A videogram vizuális objektumként való feldolgozása lehetővé teszi, hogy a beszédhangról közvetlen fizikai tapasztalattal nem rendelkező hallássérült a látáson keresztül mégis megszerezzen egy, optimális esetben a hangra megkülönböztetetten jellemző fizikai tapasztalatot. E tapasztalat birtokában és további módszertani eljárások alkalmazásával a tanuló olyan segédeszközhöz jut, ami elősegítheti számára egyrészt a beszédhangok természetes vagy ahhoz közeli kiejtését (ami a jelenlegi szurdopedagógiai gyakorlat szerint csak korlátozottan valósul meg), másrészt megtapasztalhatja a beszéd-től független környező hangzó világot is. Ezek olyan alkalmazási területek, amelyek feltétlenül érdekessé teszik az audiovizuális transzkódolás lehetőségének a vizsgálatát.

Más a helyzet a videogramok olyan feldolgozásával, ami alsóbb szintű (fonetikai és fonológiai) nyelvi feldolgozást igényel. Az ilyen feldolgozás lehetőségét kérdésessé teszi egyrészt a nyelvi és a vizuális csoportosítási modell közötti, fentebb említett lényegi eltérés, másrészt az ugyancsak említett különbség a redundanciához való viszonyban, végül nem hagyhatjuk figyelmen kívül a hallás- és a látáspercepció feldolgozási sebessége közötti jelentős kü-

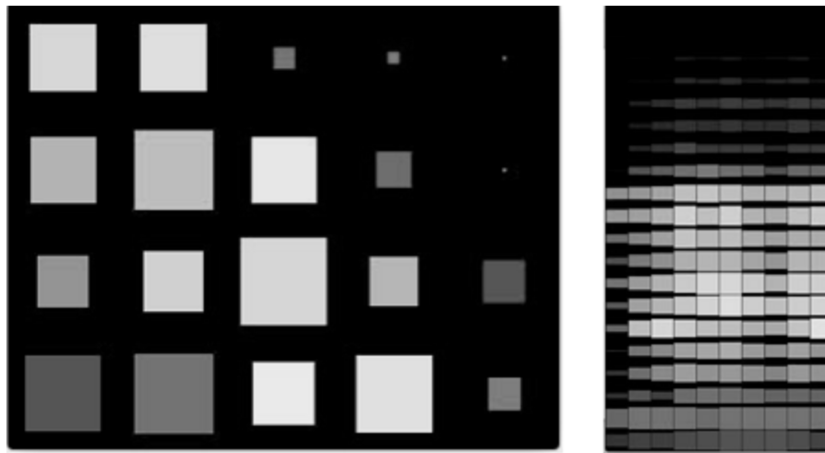
lönbséget sem. Ez utóbbi szempont akkor is lényeges lenne, ha egyébként a kétféle modalitás csoportosítási modellje és viszonyuk a redundanciához azonos lenne. Míg a beszédhangok esetében több száz, több ezer diszkrét frekvencia- és a hozzá csatlakozó intenzitásadat sikeres feldolgozásával tud a hallás megbirkózni másodpercenként, a látás esetében ez csupán 10-12 diszkrét kép. E három jelentős különbség a percepció itt tárgyalt kétféle modalitása között tehát azt sugallja, hogy az audiovizuálisan transzkódolt videogramok percepciója nem képes kellően elősegíteni a nyelvi feldolgozást. A próbálkozás kimenetét árnyalhatná ugyanakkor, ha ilyen percepciós adottságok mellett is tetten lehetne érni az agyi plaszticitás kialakulását.

Módszertan

A vizuális nyelvi feldolgozás ezen általános tulajdonságait feltárandó halló kísérleti személyekkel dolgozunk. A tesztelésben 18 halló egyetemi hallgató vesz részt (8 nő, 10 férfi magyar egyetemi hallgató, átlagéletkor a tanulás kezdetén: 21,5 év). A feladatuk három részből áll: tanulás, gyakorlás és tesztelés, melyek közül az első kettő online módon, az utóbbi megadott azonos helyszínen és időpontban történik. Ellentétben a siket gyermekekkel, akik az iskola alkalmazott pedagógiai módszerének megfelelően egész, értelmes szavak tanulásán keresztül tanulják a hangok helyes kiejtését, a halló kísérleti személyek először hangokat, majd szótagokat, végül értelmes szavakat ábrázoló videogramok felismerését sajátítják el. Annak, hogy a tanítást izolált hangok videogramjainak a felismerésével kezdtük, kettős motivációja volt: egyrészt, közvetlenül a magyar nyelvre vonatkoztatva ez tekinthető az olvasástanítás ideális sorrendjének (vö. Csépe 2006, 2014), másrészt tesztelni szeretnénk volna, mennyire alkalmas az eddig még nem alkalmazott audiovizuális transzkódolás a beszéd hangjainak az ábrázolására. Ugyanis bármennyire kézenfekvőnek tűnhet is a hang fogalma egy halló ember számára, azt a nyelvtechnológia, így az annak eszköztárára építő transzkódolásunk is a beszédfolyam pusztán fizikai jegyeiből csak igen korlátozott mértékben tudja szegmentálva felismerni. A beszédfolyamnak hangok sorozataként való értelmezése az ember egyik kimagasló kognitív teljesítménye; a technológia csupán annyit tehet, hogy az akusztikus jegyeknek egy olyan halmazát igyekszik megjeleníteni, amelynek a vizuális észlelése elősegítheti a hallók hangészlelésének alapjául is szolgáló kognitív folyamatokat, és így az agyi plaszticitás érvényesülését. Ehhez a nyelvtechnológiában használatos frekvenciatartományt, a fentebb említett Bark-skálát (Zwicker 1961; Zwicker–Terhardt 1980) használtuk: a beszédészlelésben elsődleges 125 Hz–8000 Hz-es tartomány az alapja az általunk is alkalmazott audiovizuális transzkódolásnak. Az ezen tartományon belül tovább szűrt sávok mindegyikének megfeleltettük a képernyőablak egy-egy szektorát, és az akusztikus folyam vizuális transzkódolását ezen szektorok megfelelő pozícióiban jelenítettük meg úgy, hogy ezen szektorok mérete és színe feleljen meg az eredeti akusztikus informáci-

ónak. A beszédflowam (izolált hang bemondása esetén egyetlen hang) folyamatosan, 40 ms-os mérési tartományban mért frekvencia- és intenzitásértéke-in alapuló audiovizuális transzkódolása az egyénenként és egyének között is folyamatosan változó akusztikumot tükrözi (vö. Hunyadi et al. 2014).

Kezdetben kétféle videogramot alkalmaztunk. Az egyik ún. mátrixos elrendezésben ábrázolta a bemondást oly módon, hogy egy 500×400 pixel méretű mátrix teljes felülete az adott hang 40 ms-os ablakokban mért akusztikai jellemzőit követte és jelenítette meg. A másik ábrázolás ugyanezeket az értékeket oszlopos elrendezésben jelenítette meg 600×600 pixeles méretben úgy, hogy az egyes 40 ms-onként megjelenített oszlopok egy 400 ms-os időtartamú felületen jobbról balra végigvonultak. A megjelenítésbeli különbségből adódott, hogy míg a mátrixos elrendezésben a videogram egésze 40 ms-onként folyamatosan változott (azaz „villódzott”), az oszlopos elrendezésben egy 40 ms-ra megállapított jellemző reprezentációja a jobbról balra való mozgásnak köszönhetően folyamatosan követhető volt a 400 ms-os ablak végéig. E kétféle elrendezést mutatja az 1. ábra, bal oldalon a mátrix, jobb oldalon az oszlop, mindkettő az *á* hang egy-egy momentumát reprezentálva.



1. ábra

Az *á* hang videogramja: mátrix (balra) és oszlop (jobbra)

Minden egyes bemondott hang, szótag és szó így keletkezett videogramját 18 különböző bemondóval készítettük el (a kísérleti személyek és a bemondók egyébként azonosak voltak). A tanulás során, aminek a tempóját mindenki maga választja meg, egy webes felületen egyszerre hallják az adott hangmintát és látják az annak megfelelő videogramot, így tanulva meg egy-egy videogram hangértékét. A gyakorló szakaszban (aminek az idejét ugyan-

csak szabadon megválaszthatják, figyelembe véve az egyéni előrehaladás mértékét) hang nélkül csak a videogramot látják, majd egy billentyű megnyomását követően begépelik a videogram hangji megfelelését. Lehetőség van a tanultak önellenőrzésszerű, tetszőleges számú online tesztelésére is, amikor a videogram véletlenszerű bemutatását követően a videogram hangji megfelelését gépelik be. A gép a válasz helyessége mellett méri a válaszüjtőt is. Az önellenőrzés után az alanyok százalékos értékelést kapnak a teljesítményükről, így dönthetnek a további tanulás folytatásáról. Az egyes tanítási szakaszok után (melyekben a kísérleti személyek újabb és újabb mintatípusokat sajátítanak el) egy előre meghatározott időben, zárt helyen és ellenőrzött körülmények között teszteljük a videogramok elsajátítását. A válaszok helyességének ellenőrzése mellett minden tesztben mérjük a válaszüjtőt is.

Nem a teljes magyar hangzókészletet vizsgáljuk, csupán egyes magánhangzókat, valamint olyan mássalhangzókat, amelyek a siketek beszédtanításában különösen nehezeknek bizonyulnak. A tanulás első, mintegy két hónapig tartó szakaszában a következő hangok és hangkapcsolatok videogramjai szerepeltek (a kiválasztás fontos szempontja az volt, hogy e hangok időtartamának köszönhetően hosszabb idő álljon rendelkezésre a videogramok feldolgozására): *á, i, u, v, f, zs, s, z, sz*. A második, ugyancsak két hónapos szakaszban azt vizsgáltuk, vajon mennyire tudják a kísérleti személyek elsajátítani a már tanult hangokból összeállított CV hangkapcsolatokat, azaz a lista ezeket a mintákat tartalmazta: *vá, vi, vu, fá, fi, fu, zsá, zsi, zsu, sá, si, su, zá, zi, zu, szá, szí, szu*. A harmadik szakaszban (egy hónap) arra voltunk kíváncsiak, vajon egy CV kapcsolatnál hosszabb hangzókapcsolatot is képesek-e a kísérleti személyek csupán hangok felismerése alapján elsajátítani. Itt, mint eddig is, figyelmünk a legalacsonyabb szintű nyelvi feldolgozásra irányult, azaz arra, hogy vajon megfigyelhető-e a vizuális jelsorozat hangszintű reprezentációkra való szegmentálása. A magasabb szintű, elsősorban szemantikai feldolgozást kizáró két szótagú álszavakat teszteltük: *favi, vífi, szuso, zeszé, sézsi, zsosu, vöfe, süzsa, szövi, vüszö*. Egy újabb szakaszban tovább bővítettük a tanulandó hangok listáját az *a, e, é, o, ö, ü* magánhangzókkal és az ezeket, továbbá a már korábban tanult mássalhangzókat (*v, f, zs, s, z, sz*) tartalmazó CV hangkapcsolatokkal. Az álszavak tanítását követte a jelentéssel bíró szavak tanítása (egy hónap), melyek kiválasztásában a szurdopedagógusok által pedagógiai célra összeállított százszavas szólistából válogattunk: *áru, ásó, buta, daru, fárad, gida, ír, malom, mozi, rokon, rózs, sziget, szoba, utas, váza, zene*. A kísérletek sorában utolsónak a mondatfeldolgozás szerepelt, amikor különböző szórendeket választva (íge elöl, közép, hátul) a mondat ismert szavakat tartalmazott. A szavak az eredetileg tanult, így morfoszintaktikailag nem illesztett formában szerepeltek, ezért feltettük, hogy a kísérleti személyek a szóértelmezés mellett kifejezetten szintaktikai műveleteket is megkísérlelnek mozgósítani. E tesztelést értelemszerűen erre irányuló tanulás nem előzte meg.

A mintatípus-sorozat leírásából látható, hogy a fokozatosság elvét követve először hangok reprezentációját és felismerését teszteltük azt vizsgálva, teljesíthető-e az audiovizuális transzkódolással a beszédfeldolgozás alapfeltétele, az akusztikai jelek vizuális jelekké való átalakítása (1. hipotézis), majd az egyszerű CV hangkapcsolatok, továbbá az álszavak vizsgálatán keresztül az, vajon lehetséges-e a vizuális hangszegmentálás (2. hipotézis). Végül az értelmes szavak és a mondatértés vizsgálatával arra kívántunk választ kapni, vajon az audiovizuális transzkódoláson keresztül a vizuális percepció képes-e mobilizálni a magasabb szintű nyelvi feldolgozást (3. hipotézis). A transzkódolandó hangok, hangkapcsolatok és szavak kiválasztásánál nem tudtuk maradéktalanul érvényesíteni a szigorúan vett fonetikai és fonológiai elveket (a kiválasztást sokszor szurdopedagógiai és/vagy beszédtechnológiai igények befolyásolták), azonban úgy gondoljuk, hogy e kísérletek is utalhatnak azokra a kognitív folyamatokra, amelyek a videogramok tanulása során létrejött halláspercepció/látáspercepció interfészen keresztül megvalósulnak, továbbá amelynek ismeretében a nyelvi feldolgozás egyes újabb dimenziói is feltárulhatnak.

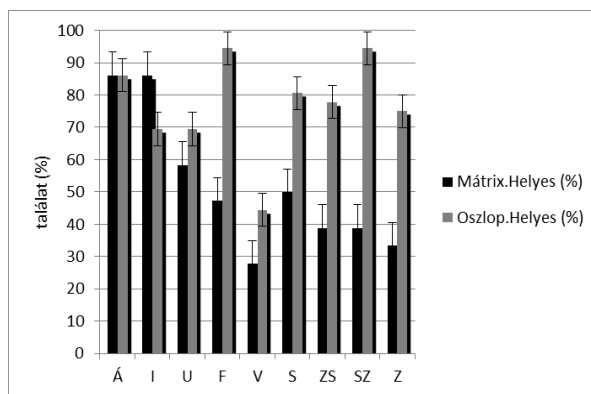
Eredmények

A videogram mint beszédhangok reprezentációja

A tanítás első szakasza egyedülálló beszédhangok vizuális ábrázolására vonatkozott. Arra voltunk kíváncsiak, lehetséges-e a beszéd akusztikai és fonetikai jegyeit úgy ábrázolni, hogy azok alapján egyes hangok (hangtípusok) egymástól megkülönböztethetően azonosíthatóak legyenek (1. 1. hipotézis). A lista a következő hangzókából állt (a rövideg-hosszúság megkülönböztetése nem volt cél, a mássalhangzóknál azonban figyeltünk a zöngés-zöngétlen párokra): *á, i, u, v, f, zs, s, z, sz*.

Ismétléses MANOVA módszerrel vizsgáltuk a hangzók esetében a videogram-mintatípusokra eső találatok közötti különbséget, amit szignifikánsnak találtunk mintatípusokon belül ($F = 0,859$; $DF = 8$; $N = 100$; $p < 0,0001$) és között ($F = 15,294$; $DF = 1$; $N = 107$; $p < 0,0001$) egyaránt. Pearson-féle χ^2 függetlenségi teszttel vizsgáltuk a válaszok alakulását a minta típusa (mátrix vagy oszlop) függvényében. Az egyedülálló hangokat tartalmazó minták tekintetében a kétféle minta között nem találtunk statisztikailag kimutatható különbséget az *á, i* és *u* magánhangzókra, valamint a *v* mássalhangzóra eső találatok alapján, ugyanakkor szignifikáns különbség ($p < 0,05$) volt az *f, s, zs, sz* és a *z* megítélésében. A mintánkénti helyes válaszokra vö. 2. ábra; válaszadók száma: 18, stimulusok száma: 54.

A CV hangzókapcsolatokat tartalmazó minták lehetőséget adtak arra, hogy a mindkét hangzó egyidejű helyes felismerése mellett teszteljük a felismerést pozíciótól (minta eleje, minta vége) függően is. Az ismétléses MANOVA-teszt szerint a videogram-mintatípusokra eső találatok közötti különbség mintatípusokon belül ($F = 2,693$; $DF = 17$; $N = 91$; $p < 0,0001$) és között ($F = 11,435$; $DF = 1$; $N = 107$; $p < 0,0001$) is szignifikáns volt.



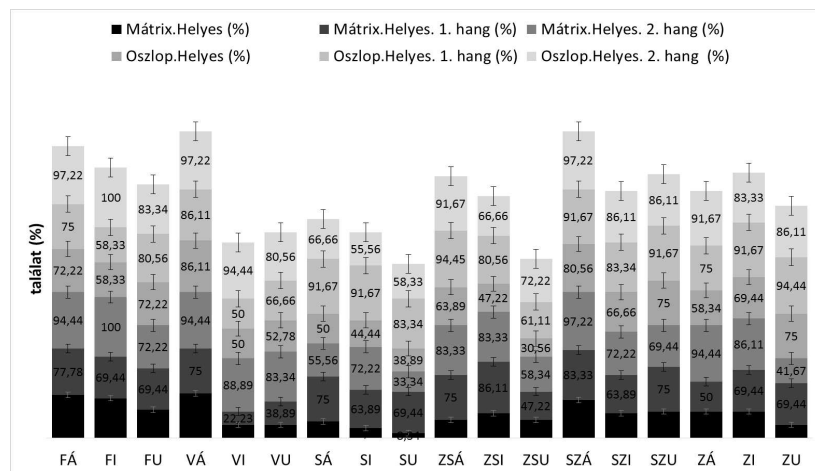
2. ábra

Egyedülálló hangzókat ábrázoló videogramok hangértékének a felismerése (mátrix és oszlop)

A mindkét elemében helyes találatok válaszadónkénti eloszlásában az *u*-t tartalmazó összes mintánál (*fu*, *vu*, *su*, *zsu*, *szu*, *zu*), továbbá a *vi* és a *zsá* esetében figyeltünk meg szignifikáns különbséget a kétféle ábrázolás között. A mintakezdő pozícióban levő hangzók felismerésekor szignifikáns különbség volt a *vi*, *vu*, *si*, *su*, *szí*, *szu*, *zi*, *zu* mátrixos és oszlopos ábrázolása között. A mintavégi pozícióban mindössze a *zu* esetében különbözött a mátrixos és az oszlopos ábrázolás szignifikánsan. A fenti adatokból arra következtetünk, hogy különösen a CV kombinációban megjelenő *u*, kisebb mértékben az *i* magánhangzó ábrázolásában/felismerésében van jelentős eltérés a kétféle megjelenítés között, és úgy, hogy az ábrázolást a koartikuláció is befolyásolja. Az, hogy ha hangkapcsolatonként különböző mértékben is, de az oszlopos ábrázolás a mátrixosnál szinte minden hangkapcsolat esetében rendre magasabb felismerési arányt mutat, sugallja az oszlopos ábrázolás előnyét (vö. 3. ábra; válaszadók száma: 18, stimulusok száma: 54).

A találatok pozíciófüggő eloszlását tekintve mátrix és oszlop között nincs szignifikáns különbség a *fá*, *fu*, *fi*, *zsá*, *zsu*, *zsi* és a *szá*, *szu*, *szí* hangsorok esetén, azaz a kétféle ábrázolás hasonló mértékben támogatja ezen hangsorok és átmenetek együttes felismerését. Ezzel szemben az ábrázolások szignifikánsan különböznek a *v* kezdőhangot tartalmazó hangsorok (*vá*, *vu*, *vi*) felismerhetősége esetén. A *v*-vel kezdődő minták esetében az első hang (a *v*) felismerése kisebb a mátrixos, mint az oszlopos mintában, ami összefügghet a mátrixos megjelenítés specifikumával. Már az ábrázolt vizuális jelek csupán rövid idejű felvillanásai miatt is az alanyok ezen hangoknak a mintakezdő első megjelenését szükségképpen nehezebben azonosítják, szemben a második hang megjelenésével, ami már egy elindult észlelési folyamat része, ahol akár a koartikulációs szakasz is „időt hagy” a hangminta felismerésére. Ez utóbbi

hatás érvényesülhet az oszlopos minták észlelésénél, ahol a 400 ms-os ablakban jobbról balra elvonuló jelek számára több idő marad a holisztikus (a környezetet is figyelembe vevő) felismerésre.



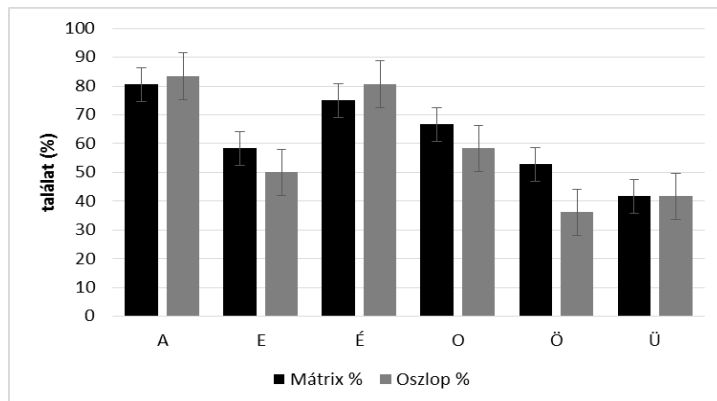
3. ábra

CV hangzókapcsolatokat ábrázoló videogramok hangértékének a felismerése (mátrix és oszlop)

Egy újabb egy hónapos tanulás során a következő magánhangzók tanítására került sor: *a, e, é, o, ö, ü* a már ismert mássalhangzókat tartalmazó újabb mintákban: *fa, fe, fé, fo, fõ, fü, va, ve, vé, vo, vö, vü, sa, se, sé, so, ső, sü, zsa, zse, zse, zso, zsö, zsü, sza, sze, szé, szo, szö, szü, za, ze, zé, zo, zö, zü* (vö. 4. és 5. ábra; válaszadók száma: 18, stimulusok száma: 54).

Az ismétléses MANOVA-teszt alapján a videogram-mintatípusokra eső találatok közötti különbség az egyedülálló magánhangzók és a CV hangzókapcsolatok esetében is szignifikáns volt magánhangzókra (F = 1,291; DF = 5; N = 31; $p < 0,0001$) és magánhangzók között (F = 8,894; DF = 1; N = 35; $p < 0,0001$), valamint CV hangzókapcsolatokra (F = 5,507; DF = 35; N = 73; $p < 0,0001$) és CV hangzókapcsolatok között (F = 7,189; DF = 1; N = 107; $p < 0,0001$). Ami a mátrixos és az oszlopos ábrázolás megkülönböztetését illeti, ez Pearson-féle χ^2 -teszt szerint az egyedülálló hangokat tekintve csak az *a* esetében volt szignifikáns. Ami a CV kapcsolatot illeti, bár szinte minden mintában abszolút értékekkel gyengébben szerepelt a mátrix (a mátrixos ábrázolás eredményezte a legtöbb olyan felismerést, ahol az alanyok egyik hangot sem tudták helyesen azonosítani), a mátrix és az oszlop összevetése összességében csak négy mintában volt szignifikáns: a *fa* és a *fe* esetében a mátrix, a *su* és a *sü* esetében az oszlop javára. A minta kezdő pozíciójában levő hang fel-

ismerése a minták alig több mint egyharmadában különbözik szignifikánsan a mátrix és az oszlop mintatípusok szerint: a mátrix javára a *fa*, *fe*, *fé* kapcsolatban, az oszlop javára a *va*, *ve*, *vé*, *vo*, *sa*, *zsa*, *zse*, *zsé*, *zso*, *zsö*, *szö* és *zü* mintában. A minta végső pozíciójában a mátrix és az oszlop megkülönböztetése egyetlen összevetésben sem volt szignifikáns.



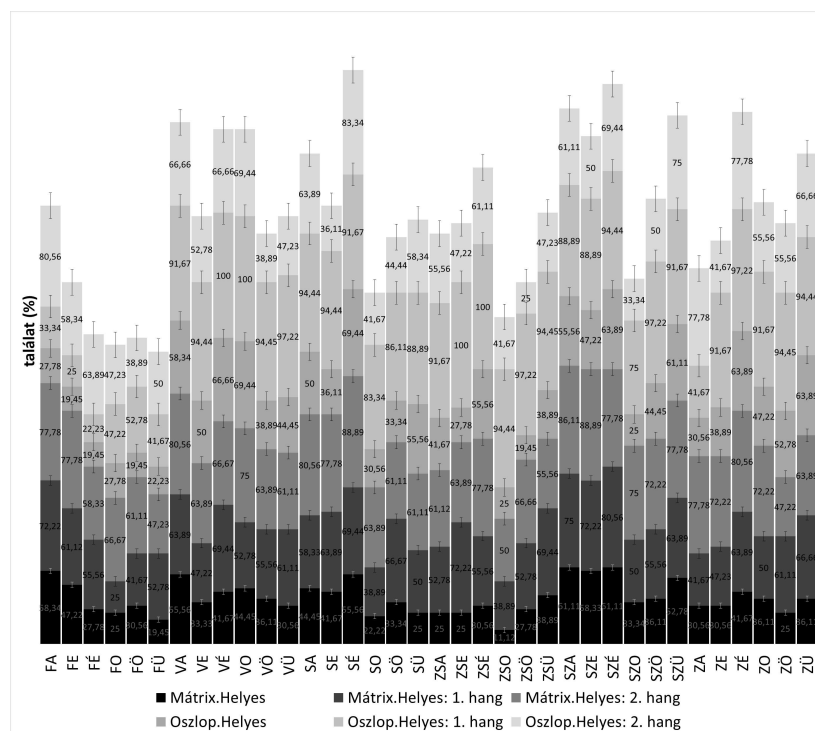
4. ábra

Egyedülálló hangzókat tartalmazó videogramok hangértékének a felismerése.
2. sorozat (mátrix és oszlop)

Hasonlóan az első teszt sorozat elemzéséhez, ezen újabb adatok alapján is feltehetjük, hogy a mátrix gyengébb teljesítményéhez nem elhanyagolható mértékben hozzájárul maga a minta prezentálásának a módja: az, hogy mire a vizuális feldolgozás interpretálhatná az észlelt mintát, a videogram a 40 ms-os változó ablaka miatt az észlelés számára már nem elérhető. Ugyanakkor az ezt követő, második pozícióban már nem találunk szignifikáns különbséget a mátrix és az oszlop között egyetlen minta esetében sem. Ez utalhat arra, hogy a már korábban aktiválódott percepció folyamatában a kísérleti személy érzékelheti (és akár feldolgozhatja) a második hang teljes mintázatát is. A tapasztalt nagyobb mintavégi találati arány is közvetve utalhat erre a késleltett percepcióra, ami valóban független lehet a megjelenítés típusától.

Annak érdekében, hogy megválaszoljuk, vajon egyszerű vizuális alakfelismerés eredménye-e a hangfelismerés, vagy eközben megvalósul-e fonetikai szegmentálás is, olyan hangkapcsolatok felismerését is teszteltünk – külön tanítási szakasz nélkül –, amelyeket bár teljes mintaként nem tanultak, egyes összetevőiket külön-külön igen. Ezek olyan hangkapcsolatok voltak, amelyek szerkezete CVCV volt, és amelyek CV összetevőit korábban már külön tanulták. A fonetikai szegmentálás ellenőrzését elősegítette, egyben a felismerést nehezítette, hogy ezek jelentés nélküli, továbbá a fonotaktikának is kevésbé

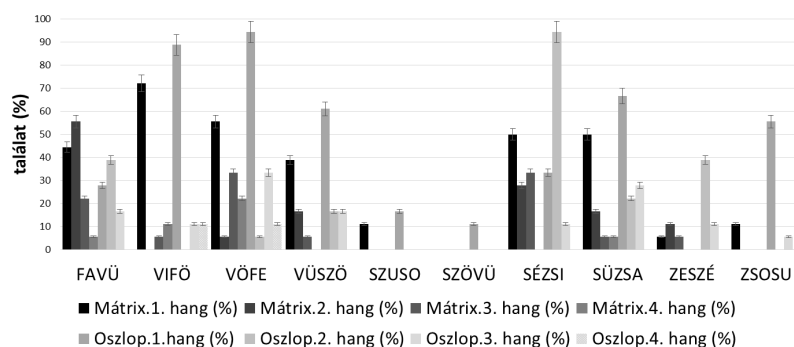
megfelelő álszavak voltak: *favü, vífö, szuso, zeszé, sézsi, zsosu, vöfe, süzsa, szövü, vüszö* (vö. 6. ábra; válaszadók száma: 18, stimulusok száma: 20).



5. ábra

CV hangzókapcsolatokat ábrázoló videogramok hangértékének a felismerése (mátrix és oszlop, második sorozat)

Az ismétléses MANOVA-teszt alapján a videogram-mintatípusokra eső találatok közötti különbség az álszavak esetében is szignifikáns volt mintatípusokon belül ($F = 0,982$; $DF = 9$; $N = 171$; $p < 0,0001$) és között ($F = 1,021$; $DF = 1$; $N = 179$; $p < 0,0001$) egyaránt. Figyelemre méltó, hogy egyetlen álszó esetében sem történt teljes felismerés. A részleges felismerések tekintetében a mátrix- és az oszlopmintázat eloszlása között csak a *szövü* álszó esetében volt szignifikáns különbség a válaszokban (Pearson-féle χ^2 : $p = 0,043$). Ami pedig a felismerendő hangok pozícióját illeti, az itt is a többitől eltérően viselkedő *szövü* álszó kivételével (amelyre összesen három helyes találat esett, egy az első, kettő a második hangra és a továbbiakra egy sem) az összes álszóra eső találatok szignifikánsan pozíciófüggőek voltak ($p \leq 0,003$).



6. ábra

Álszavakat ábrázoló videogramok hangértékének a felismerése (mátrix és oszlop)

A minden hangra, valamint a harmadik és a negyedik hangra első helyes találatok hiánya, továbbá az a tény, hogy a legtöbb helyes találat a második hangra esett, megerősíti a korábbi érvelésünket: a vizuális percepció mintatípustól függetlenül nem tudja feldolgozni a két hangnál hosszabb mintákat, továbbá a mintakezdő hang felismerése sokkal kevésbé valószínű, mint a másodiké. Az a tény, hogy a korábban tanult CV hangkapcsolatok felismerése akár a minta kezdetén is problematikus volt, azt jelzi, hogy bár a kísérleti személyek korábban megtanulhatták az adott CV komponensek „jelentését”, azonban e tanult mintát nem tudták azonosítani mint a teljes álszó egy „értelmes” összetevőjét. Ennek a feldolgozás sebességén túlmenően oka lehet a szegmentálás hiánya is.

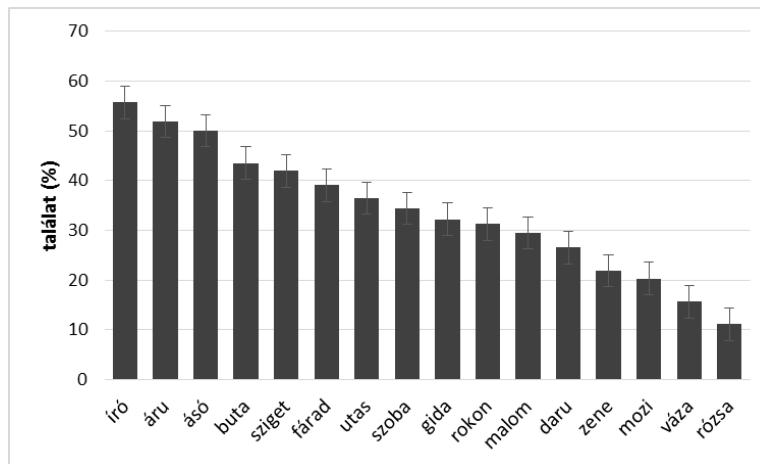
A videogram és a magasabb szintű, szemantikai nyelvi feldolgozás

A kutatás természetes folytatásaként megkíséreltük értelmes szavak tanítását is, újabb mássalhangzókkal bővítve a hangkészletet. Az egy hónapig tartó tanítás arra a vizuális beszédfeldolgozás szempontjából igen fontos kérdésre irányult, vajon társul-e a vizuális alakfelismerés szemantikai interpretációval, és ez vajon hatékonyabbá teszi-e az alakfelismerést. Itt és a továbbiakban csak a sikeresebb felismerést biztosító oszlopos ábrázolást alkalmaztunk. Az eredményeket a 7. ábra mutatja (válaszadók száma: 18, stimulusok száma: 54).

Az ismétléses MANOVA-teszt alapján a videogram-mintatípusokra első találatok közötti különbség az értelmes szavak esetében szignifikáns volt mintatípusokon belül ($F = 4,201$; $DF = 15$; $N = 14$; $p = 0,007$) és között ($F = 4,229$; $DF = 1$; $N = 28$; $p < 0,0001$) egyaránt.

Mint az ábrából is kitűnik, az álszavak felismerésével összevetve (amelyeknél egyetlen teljes szó felismerése sem volt sikeres), most több értelmes szó teljes felismerése is elég magas arányt ért el. Ez az eredmény pozitív módon meg is válaszolja a kísérletben feltett kérdést: a vizuális minták, ha sze-

mantikai tartalom köthető hozzájuk, sikerebben felismerhetők, mint ha szemantikailag értelmezhetetlenek (hasonlóan ahhoz, ahogyan egy jelentéssel bíró szót könnyebb feldolgoznunk, mint amihez jelentést nem tudunk társítani; vö. Wheeler 1970; Lukács et al. 2014).



7. ábra
Értelmes szavak tanulása (oszlop)

A videogram és a magasabb szintű, szintaktikai nyelvi feldolgozás

Kísérleteink utolsó szakaszában (újabb egy hónapig) folytattuk további értelmes szavak tanítását, a cél a szófelismerésen kívül a vizuális mondatfeldolgozás/-értés vizsgálata volt, újra csak oszlopos ábrázolásban. Bár a kísérlet mondatfeldolgozásra/-értésre irányult, a kísérleti személyek a szavakat nem mondatokban, hanem izoláltan tanulták, így a tesztben stimulusként szereplő mondatokat korábban nem látták. A szavak (melyeket a szurdopedagógusok által javasolt, pedagógiai szempontból kiemelt százas szólistájából válogattuk) a következők voltak: (alanyi szerepre szánva:) *áru, ásó, buta, daru, gida, író, malom, mozi, rokon, rózsza, sziget, szoba, utas, váza, zene*, (állítmányi szerepre szánva:) *akar, csinos, éhes, eladó, fárad, nyal, okos, pici, pihen, sós, utál, üzen*. E szavakból igyekeztünk értelmes mondatokat összefűzni, és a tanítási szakaszt követően e mondatokban a felismerésüket tesztelni. (Bevalljuk, az így kapott értelem helyenként nehézkesre vagy akár mulatságosra is sikeredett.) A mondatok a következők voltak (mondatonként három különböző szórenddel; a szavakat ragozás nélkül adtuk meg, ahogyan izoláltan tanulták): *Éhes gida nyal rózsza, Pici ásó eladó, Buta író üzen utas, Éhes író fárad, Éhes utas fárad, Csinos rokon ír mozi, Sós áru utas utál, Fárad utas pihen szoba, Okos gida akar pihen, Csinos rokon akar rózsza*.

(Mindhárom mondatváltozatot figyelembe véve a megjelenített összes szó száma 103 volt.)

Tapasztalataink meglehetősen vegyesek voltak. Egyrészt az alanyok körülbelül fele képes volt egynél több szót is felismerni a 3–5 szóból álló mondatokban (egyharmaduk három szót is). A mondat szavakra való szegmentálása hasonló elvet követhetett, mint amit a hangzó beszédben lehet tapasztalni (vö. Lukács 2014): a tanult/gyakorolt vizuális szóminták belsejében tapasztalt mintaátmenetek gyakoriságához képest a nem tanult szóátmenetek gyakorisága jelentősen kisebb lehetett, így azokat nem tekintették valamely adott szóhoz tartozónak. Másrészt vártuk, hogy a csak egy morfológiai alakban tanult szóalakok sorozatát mondatként értelmezve a mondatot morfoszintaktikailag helyesen (tehát továbbragozva) fogják visszaadni. Az eredmények szerint ilyen ragozott mondatvisszaadás a 15 alany közül csak három esetében történt (leszámítva azokat az eseteket, amikor a mondat helyes visszaadásához a szavakat továbbragozni nem kellett).

A mondatértékes tesztben szereplő 15 kísérleti személyből legtöbben, 11-en hibátlanul az *Eladó pici ásó* mondatot találták el. Ezt követte 6-6 találattal a *Fárad éhes utas* és a *Pici ásó eladó*. Hárman értelmezték e szósort: *Éhes nyal rózsá az Éhes gida nyal rózsá* „mondatból”, és ugyancsak hárman az *Éhes utas fárad(t)* mondat mindhárom szavát. A *fárad* szót bizonyos szövegkörnyezetben *fáradt* alakban adták vissza, mintegy követve a mondatként észlelt szósorozat értelmét is. A *fáradt* szóalak morfoszintaktikai illesztése egyébként külön figyelemre méltó. Úgy tűnik, hogy a mintaként tanult *fárad* alakot nem egy egyszerű (jelen idő, egyes szám, harmadik személyű) szóalaknak tekintették, hanem egy lexéma prototipikus megnevezésének, ami a vizsgálatunk tárgyát képező vizuális motoros reprezentációban nyelvi feldolgozáson ment át egy, a mentális lexikontól független mentális nyelvtanra támaszkodva (ugyancsak e kettő megkülönböztetését tételezi fel hangzó nyelv alapján Pinker 1991; Clahsen 1999).

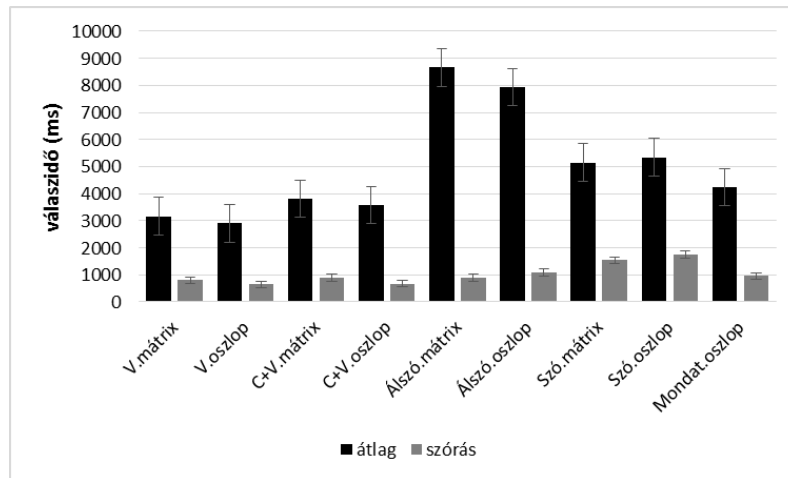
A feldolgozás sebessége

Végül, miután számos egyéb szempontot megvizsgáltunk, nem hagyhatjuk figyelmen kívül a feldolgozás sebességét. Mivel a válaszütem utal a feldolgozás sebességére, ennek alapján közvetve utalást nyerhetünk arra, mennyire bizonyultak egyes mintatípusok nehezebbeknek vagy könnyebbeknek az ábrázolás módjának (mátrix vagy oszlop) függvényében (vö. 8. ábra).

Az ismétléses ANOVA szerint a mátrixos és az oszlopos ábrázoláshoz fűződő válaszütemek szignifikánsan különböztek mind alanyonként ($F = 9,163$; $DF = 8$; $N = 9$; $p = 0,001$), mind alanyok között ($F = 18,216$; $DF = 1$; $N = 16$; $p < 0,0001$).

Ami az ábrázolás módját illeti, az egyes mintatípusokon (pl. egyedülálló magánhangzó, álszó, értelmes szó stb.) belül a mátrix és az oszlop között a válaszütem alapján nem találtunk szignifikáns különbséget. Ugyancsak figyelemre méltó, hogy az egyetlen magánhangzót, valamint a CV kapcsolatot tar-

talmazó minták reakcióideje sem különült el egymástól. Ez sugallhatja azt, hogy az interpretáció hangisegmentálás nélkül történt. Az, hogy a fenti mintákban a helyes és helytelen válaszok a reakcióidő alapján nem különültek el, újra csak azt a véleményünket erősítheti, hogy a felismeréssel nem járt együtt valamiféle nyelvi (itt: szemantikai) feldolgozás.



8. ábra
Az átlagos válaszidők mintatípusonként

A nyelvfeldolgozás szempontjából lényeges kérdés, vajon megjelenik-e szemantikai feldolgozás az értelmes szavak videogramjainak az értelmezése során. Az álszavak és értelmes szavak összevetésétől azt várjuk ugyanis, hogy más kognitív mechanizmus játszódik le akkor, amikor értelmes szavakat ábrázoló mintát kell felismernünk (a vizuális minta felismeréséhez ekkor mozgósítani tudjuk az adott alakzathoz kötött szemantikát), mint akkor, amikor csupán a vizuális jelekre mint geometriai konfigurációra (alakzatra) támaszkodhatunk. Azt is sejthetjük, hogy a többféle modalitás szinkrón alkalmazásával a percepció értelmezése nemcsak sikeresebb, de egyben gyorsabb is lesz. A sikeresebb értelmezés jól szemléltethető a fentebbi 5. és a 6. ábra összevetésében (az álszavak esetében teljes „szó” értelmezése nem is valósult meg), a 7. ábra pedig jól példázza az értelmes szavak felismeréséhez kapcsolódó jelentősen rövidebb válaszidőt. Az ismétléses MANOVA alapján szignifikáns különbséget találtunk az álszavak, értelmes szavak és ismeretlen mondatból felismert szavak válaszüzeje alapján mind a válaszadókon belül ($F = 2,543$; $DF = 4$; $N = 13$; $p = 0,0015$), mind azok között ($F = 13,334$; $DF = 1$; $N = 16$; $p < 0,0001$), miközben a Spearman-féle χ^2 -teszt nem talált szignifikáns különbséget mátrix és oszlop között egyik szótípusban sem. Ennek alapján

feltesszük, hogy a válaszdőkben megmutatkozó szignifikáns különbség a nyelvi feldolgozásnak és az abban rejlő különbségeknek (szemantikai értelmezhetőség és ilyen értelmezés hiánya) tudható be.

Abból, hogy a CV hangkapcsolat felismerése relatíve kisebb különbségű válaszdő-növekedéssel jár a magánhangzók felismeréséhez képest, mint amennyivel azoknál több hangból áll, továbbá, hogy az értelmes szavak felismeréséhez kapcsolódó válaszdő relatíve kevésbé növekszik a fele annyi jelből álló CV hangkapcsolatok válaszidejéhez képest, azt sugallja, hogy a videogramok értelmezésének a percepció alapja nem lehet a hangokra történő szegmentálás. Kézenfekvőnek tűnik azt gondolni, hogy a videogramok feldolgozása során a szegmentálásban testet öltő alsóbb szintű nyelvi feldolgozás helyett egy holisztikus, Gestalt-jellegű vizuális alakértelmezéssel lehet dolgunk. A leghosszabb jelsorozatból a többszavas mondatok állnak, miközben a válaszdő éppen ebben a mintatípusban a relatíve legrövidebb (abszolút értelemben még rövidebb is a 3-5-ször kevesebb jelet tartalmazó önálló értelmes szavak válaszidejénél). Ez a mintatípus mutatja leginkább a jelentőségét a percepció multimodális értelmezésének: a vizuális mintaillesztéshez társul a nyelvi feldolgozás két fontos szintje, a szemantikai és a szintaktikai. Úgy tűnik, hogy a nyelvi feldolgozás valóban gyorsabb, mint a vizuális alakfelismerés, ráadásul az egy időben két szinten is zajló nyelvi feldolgozás magát a nyelvi feldolgozást is gyorsítja (mint láttuk, a „csak” egyféle – szemantikai – nyelvi feldolgozásnak helyt adó, izolált szó szintű percepció és értelmezés abszolút értelemben még akár hosszabb időt is vesz igénybe, mint a „bonyolultabb”, szintaktikai feldolgozást is magába foglaló mondatértés). Ha a fentiek alapján helyes lehet az az általánosításunk, hogy minél több nyelvi modalitást vesz igénybe a percepció és az értelmezés, annál gyorsabban lesznek azok eredményesek, akkor megfogalmazhatjuk azt a feltevést, hogy a nyelvi feldolgozás a különböző szinteken egy időben, párhuzamosan megy végbe (ellentétben Friederici szintaxisközpontú és elkülönült, egymásra épülő fázisokat feltételező modelljével, vö. Friederici 1999, 2002). A kísérleteink tárgyát képező, a nyelvvel kisebb-nagyobb mértékben kapcsolatba hozható vizuális minták tanulásával és felismerésével kapcsolatos megfigyeléseink legalábbis ezt sugallják; a nyelvi feldolgozás ilyen értelmezése egyébként illeszkedik Hagoort (2005) információkat egyesítő unifikációs modelljéhez és van der Helmnek (2014) a perceptuális szervezés transzpárhuzamos voltára vonatkozó javaslatához. Mindezek mellett figyelembe kell venni azt, hogy adatainkban igen hosszú, 2500-8000 ms-os átlagos válaszdőkről van szó. E tény lényegesen befolyásolni fogja érvelésünket, amikor azt a kérdést vizsgáljuk, mennyiben alkalmazható az audiovizuális transzkódolás mint a valós idejű beszéd vizuális reprezentációjának módszere.

Összefoglalás

Az eredményeket összefoglalva, és azokat a kiinduló hipotézisekre kivetítve a következőket állapíthatjuk meg. A kísérlet elméleti középpontjában há-

rom hipotézis ellenőrzése állt, és ennek céljából percepciók kísérleteket végeztünk.

Az 1. hipotézisben azt tettük fel, hogy az audiovizuális transzkódolás alkalmas arra, hogy a videogramokon keresztül a beszéd fonetikai jegyeit szükséges és elégséges módon ábrázolja. E hipotézis csak részben teljesült: egyes hangok, különösen az *a* és az *i* magánhangzók, továbbá egyes mássalhangzók, így az *sz* és a *f* izoláltan jól felismerhetők voltak. A zöngésséget/zöngétlenséget mint fonetikai jegyet ugyancsak megfelelően felismerték. Ezzel szemben a CV hangzókapcsolatban gyakran nem ismertek fel egyes kezdő mássalhangzókat, mint pl. a *v*-t. A álszavak vizsgálatából az is kiderült, hogy egy és ugyanazon hang első szótagbeli előfordulását sikeresebben felismerték, mint amikor az a második szótagban fordult elő. Ez a megfigyelés elvezet bennünket a 2. hipotézis vizsgálatához.

A 2. hipotézis a vizuális reprezentáció és az alsóbb szintű nyelvi feldolgozás kapcsolatáról szólt és feltette, hogy a beszéd videogramokon keresztüli reprezentációja lehetővé teszi, hogy e reprezentációban beszédhangokat szegmentáljunk. Mint láttuk, egyes hangokat izoláltan sem sikerült kellően felismerni; ennek okát lehetett akár az ábrázolás alapjául szolgáló eljárásban is keresni. A CVCV szerkezetű álszavakat a maguk teljességében nem sikerült felismerni a korábban tanult, ugyanezen CV kapcsolatokat tartalmazó videogram-mintázatokból, azaz nem szegmentálták e sort két már tanult CV kapcsolatra. Ennek oka lehet a feldolgozás időigényes volta is. Azonban az, hogy a bemutatott álszavak szótagszámát (minden minta két szótagból állt) szinte kivétel nélkül jól érzékelték még akkor is, ha az adott szótagok magánhangzóit csak több-kevesebb sikerrel találták el, azt mutatja, hogy a hangzó beszéd szempontjából lényeges szótagfelismerés (a videogram szótagra történő szegmentálása) megtörténik. Mivel ugyancsak a legtöbb esetben a magánhangzó előtt észleltek egy mássalhangzót is, arra kell következtetnünk, hogy a kísérleti személyeknek sikerült egy CV szerkezetű hangsor videogramját egy mássalhangzóra és egy magánhangzóra, és ebben a sorrendben szegmentálni. Így a hipotézissel kapcsolatban azt összegezhethetjük, hogy úgy tűnik, bár igen korlátozottan, de megvalósul a videogramok alsóbb szintű, fonetikai szegmentálása magánhangzókra és mássalhangzókra. Ugyanakkor a kérdés nyitva marad, mivel magyarázzuk a specifikus, hangszintű felismerés jóval kisebb arányát. Ennek okát kereshetjük akár az audiovizuális leképezés nem elégséges voltában, akár a feldolgozás sebességi korlátaiban, akár egyidejűleg mindkét tényezőben.

A 3. hipotézis szerint az audiovizuális transzkódolás lehetővé teszi a vizuális alapú magasabb szintű nyelvi feldolgozást, beleértve a fonémareprezentációt, a szóazonosítást és a mondatértést. Ezen hipotézis valójában a vizuális jelnek nyelvhez köthető, de a jelentést megelőző absztrakt grammatikai, továbbá szemantikai értelmezéséről szól. Bár az egyes szavak felismerése alanyonként jelentősen különbözött, meg tapasztalhattuk a szintaktikai feldolgo-

zás jelenlétét: az egymást követő szavak „összeolvasását” egyetlen értelmes mondatba. Abból, hogy a *Fárad éhes utas* mondat különböző változataihoz hasonló mértékben értelmeztek olyat, amiben a konkrét mintaként tanult *fárad* helyett a *fáradt* alak szerepel (*fáradt éhes utas*, *fáradt és éhes utas*, *fáradt az éhes író*, *fáradt éhes az író*), túl azon, hogy működni látszik mind a szószemantika, mind a morfoszintaxis és a szintaxis, felvetjük, hogy valamilyen fonológiai értelmezés is megvalósulhatott, azaz a hangzó beszédben megkülönböztetett *fárad* és *fáradt* formáknak egyaránt megfeleltették a prototipikusan *fáradnak* tanult szóalakot, mintegy fonematikusan értelmezve azt. A fonetikai elemzés is tetten érhető: egyes esetekben az alanyok fonetikai asszociációs stratégiát (vö. Gósy 1999, 2000) alkalmaztak a videogramok szóértelmezésében: háromszor az *utas* helyett *úrt* „olvastak” (*Fárad éhes utas* > *Fárad éhes úr*), egyszer pedig az *író* helyett ugyancsak *úr* jelenik meg (*Éhes úr fárad*).

Következtetések

A halló kísérleti személyekkel végzett tesztek eredményei azt mutatják, hogy a beszéd akusztikus jelfolyamának vizuális transzkódolásával létrehozott ábrázolásokat a vizuális percepció segítségével komoly megszorítások mentén képesek vagyunk hangokként és hangok sorozataként értelmezni. További kísérletekre lenne azonban szükség, hogy megtudhassuk, kialakítható-e a hang és a fonéma megkülönböztetése a hallássérülteknél (így pl. értelmezi-e a mentális reprezentáció a *fáradt*, továbbá a *fáradj* videogramját úgy, mint ami a *fárad* fonematikus prototípus egy-egy változata). Kitűnt, hogy – jelentős egyéni különbségek mellett – a videogramok percepciója lehetővé teszi a szemantikai, továbbá a morfoszintaktikai és szintaktikai nyelvi feldolgozást, és – összhangban az olvasáskutatás sokoldalúan igazolt eredményeivel – úgy találtuk, hogy a minták többszintű nyelvi értelmezése előnyös azok sikeres felismerése szempontjából. Mindezek alapján megfogalmaztuk a különböző szintű nyelvi feldolgozások párhuzamosságának a feltevését.

Azt, hogy a transzkódolás alkalmas lehet a folyó beszéd valós időben történő vizuális reprezentálására és értelmezésére is, a percepciós kísérletek nem támasztották alá. Ennek okai között neveztük meg a hallás- és a látáspercepció közötti lényeges különbségeket: a két modalitás modalitásspecifikus objektumai csoportosításának eltérő modelljei, egymásnak ellentmondó viszonyuk a redundanciához, végül eltérő feldolgozási sebességük. Ismét hangsúlyoznunk kell ugyanakkor, hogy e megállapításokat halló személyekkel végzett kísérletek alapján vontuk le. Mivel a felnőttekhez hasonló percepciós kísérleteket hallássérült gyerekeknél nem végeztünk, nem vehettük figyelembe az életkori sajátságokat. Nem teszteltük a vizuális figyelem mentén megfigyelhető, az olvasást befolyásoló jellemzőket (vö. Facoetti et al. 2010), amelyek feltehetően a videogramok feldolgozását is befolyásolják. Sem azt, hogy az olvasás egyre jobb elsajátításával a vizuális feldolgozás bizonyos tí-

pusai sérülnek (vö. Pegado et al. 2014), így elképzelhető, hogy a videogramok egy korábbi elsajátítása esetén a nyelvi feldolgozás is sikeresebb lehet. Ugyancsak nem vettük figyelembe a hallássérülteknek, különösen a születetten siketeknek a hallókétől feltételezhetően eltérő agyi struktúráját, ami akár a vizuális hangfeldolgozás képességét is kibontakoztathatja. Végül várjuk azon képalkotó vizsgálatok eredményeit, amelyek megmutathatják, volt-e hatása a videogramok tanításának az agyi plaszticitásra, a modalitásközi megfeleltetésen és integráción alapuló percepcióban részt vevő egyes agyi területek átstrukturálásában. Ha ilyen hatás kimutatható lenne (de csak ekkor!), az további koncentrált tanítással, az olvasási fluencia kialakításával rövidebbé, praktikusán elfogadhatóbbá tehetné a jelen adatok alapján irreálisan hosszú feldolgozási sebességet is.

Irodalom

- Bartha Csilla – Hattyár Helga – Szabó Mária Helga 2006. A magyarországi siketek közössége és a magyarországi jelnyelv. In Kiefer Ferenc – Siptár Péter (szerk.): *Magyar nyelv*. Akadémiai Kiadó, Budapest. 852–906.
- Clahsen, Harald 1999. Lexical entries and rules of language: A multidisciplinary study of German inflections. *Behavioral and Brain Sciences* 22. 991–1013.
- Csépe Valéria 2006. *Az olvasó agy*. Akadémiai Kiadó, Budapest.
- Csépe Valéria 2014. Az olvasás rendszere, fejlődése és modelljei. In Pléh Csaba – Lukács Ágnes (szerk.): *Pszicholingvisztika. Magyar pszicholingvisztikai kézikönyv 1*. Akadémiai Kiadó, Budapest. 339–370.
- Facoetti, Andrea – Corradi, Nicora – Ruffino, Milena – Gori, Simone – Zorzi, Marco 2010. Visual spatial attention and speech segmentation are both impaired in preschoolers at familial risk for developmental dyslexia. *Dyslexia* 16/3. 226–223.
- Gósy Mária 1999. *Pszicholingvisztika*. Corvina, Budapest.
- Gósy Mária 2000. Az életkor hatása a mentális lexikon működésére. *Magyar Nyelvőr* 124/4. 410–423.
- van der Helm, Peter A. 2014. *Simplicity in vision. A multidisciplinary account of perceptual organization*. Cambridge University Press, Cambridge.
- Hattyár Helga 2009. A siketközösségek kialakulása és néhány főbb jellegzetessége. In Laki Ildikó (szerk.): *Fogyatékoság és a mai magyar társadalom*. Belvedere Meridionale Alapítvány, MTA Szociológiai Kutatóintézet, Budapest, Szeged. 69–83.
- Hickok, Gregory – Poeppel, David 2004. Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition* 92. 67–99.
- Hickok, Gregory – Poeppel, David 2007. The cortical organization of speech processing. *Nature Reviews Neuroscience* 8/5. 393–402.
- Hickok, Gregory – Houde, John – Rong, Feng 2011. Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron* 69. 407–422.
- Hunyadi, László – Szekrényes, István – Czap, László – Sziklai, István 2014. Seeing the sounds? *Argumentum* 10. 325–338.

- Lukács Ágnes 2014. Szótanulás. In Pléh Csaba – Lukács Ágnes (szerk.): *Pszicholingvisztika. Magyar pszicholingvisztikai kézikönyv 1.* Akadémiai Kiadó, Budapest. 521–549.
- Lukács Ágnes – Pléh Csaba – Kas Bence – Thuma Orsolya 2014. A szavak mentális reprezentációja és az alaktani feldolgozás. In Pléh Csaba – Lukács Ágnes (szerk.): *Pszicholingvisztika. Magyar pszicholingvisztikai kézikönyv 1.* Akadémiai Kiadó, Budapest. 167–250.
- von Muenster, Kylie – Baker, Elise 2014. Oral communicating children using a cochlear implant: Good reading outcomes are linked to better language and phonological processing abilities. *International Journal of Pediatric Otorhinolaryngology* 78/3. 433–444.
- Näätänen, Risto – Lehtokoski, Anne – Lennes, Mietta – Cheour, Marie – Huotilainen, Minna – Iivonen, Antti – Vainio, Martti – Alku, Paavo – Ilmoniemi, Risto J. – Luuk, Aavo – Allik, Jüri – Sinkkonen, Janne – Alho, Kimmo 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385. 423–434.
- Pegado, Felipe – Comerlato, Enio – Ventura, Fabricio – Jobert, Antoinette – Nakamura, Kimihiro – Buiatti, Marco – Ventura, Paolo – Dehaene-Lambertz, Ghislaine – Kolinsky, Régine – Moraish, José – Braga, Lucia W. – Cohen, Laurent – Dehaene, Stanislas 2014. Timing the impact of literacy on visual processing. In: *Proceedings of the National Academy of Sciences* 111/49. 5233–5242.
- Pinker, Steven 1991. Rules of language. *Science* 253. 530–535.
- Ramus, Franck – Rosen, Stuart – Dakin, Steven C. – Day, Brian L. – Castellote, Juan M. – White, Sarah – Frith, Uta 2003. Theories of developmental dyslexia: Insights from a multiple case study of dyslexic adults. *Brain* 126/4. 841–865.
- Wertheimer, Max 1924. Über Gestalttheorie. *Symposion* 1. 39–60.
- Wertheimer, Max 1925. *Drei Abhandlungen zur Gestalttheorie*. New edition. Verlag der Philosophischen Akademie, Erlangen.
- Wheeler, Daniel 1970. Processes in word recognition. *Cognitive Psychology* 1. 59–85.
- Zwicker, Eberhard 1961. Subdivision of the audible frequency range into critical bands. *Journal of the Acoustical Society of America* 33/2. 248–248.
- Zwicker, Eberhard – Terhardt, Ernst 1980. Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *Journal of the Acoustical Society of America* 68/5. 1523–1525.

Köszönjük a TÁMOP 4.2.2.C.-11/1/KONV-2012-0002 pályázatnak a jelen tanulmány létrejöttéhez nyújtott támogatását. Az alapkoncepció kidolgozója Sziklai István (Debreceni Egyetem, Fül-, Orr-, Gége- és Nyaksebészeti Klinika), az audiovizuális transzkódolás kidolgozója és megvalósítója Czap László (Miskolci Egyetem, Automatizálási és Infokommunikációs Tanszék), a videogramok percepció relevanciáját Hunyadi László és Szekrényes István (Debreceni Egyetem, Általános és Alkalmazott Nyelvészeti Tanszék) vizsgálta. A hallássérültek tanítása hallássérültek három iskolájában (Debrecen, Eger, Budapest) jelenleg is folyik.

EGY HIÁNYPÓTLÓ ADATBÁZIS: A TINI BEA

Gyarmathy Dorottya – Neuberger Tilda

Bevezetés

A beszéd kialakulásához szükséges többek között az artikulációs mozgások állandósulása, a mentális lexikon elemekkel való feltöltése, az anyanyelvi fonológiai és grammatikai szabályok, illetve a pragmatikai sajátosságok elsajátítása. Ahhoz, hogy egy gyermek képessé váljon jelentéssel bíró, jól formált közléseket létrehozni, több területet érintő fejlődésre van szüksége, például a fonológia, lexika, morfológia, szintaxis és pragmatika terén (pl. Ninio–Snow 1996; Vihman 1996). A beszéd megindulásának alapvető feltétele a jó hallás, a pontos észlelés és a megfelelő beszédfeldolgozás. A gyermeknyelvi kutatások többsége a produkciós mechanizmus leírására fektet nagyobb hangsúlyt (különösen a gyermekek első éveire vonatkozóan), és kevesebb figyelem irányul a beszédfeldolgozási működésekre. Ennek feltehetően az a legfőbb oka, hogy a produkció jelenségei közvetlenül tanulmányozhatók, míg a percepció működése rejtetten zajlanak, megismerésükhöz célzott kísérletek szükségesek, amelyek kivitelezése a gyermek adatközlőkkel nehezebb, olykor gyakorlatilag megoldhatatlan. Kísérleti helyzetben a kisgyermekek – a felnőttekhez képest is nagyobb mértékben – kényelmetlenül, feszélyezetten érezhetik magukat, ezért kevésbé kooperatívak a (sokszor számukra ismeretlen) kísérletvezető személlyel.

A gyermekek beszédprodukciós teljesítménye sokféle módszerrel vizsgálható. Megbecsülhető például tesztelés nélkül, hétköznapi mintavételezéssel, hiszen egy felnőttel vagy egy másik gyermekkel folytatott párbeszéd sok információt szolgáltat a kifejezőképességükről. A szülői napló a gyermeknyelv megfigyelésének legegyszerűbb és legrégebbi módszere (Mervis et al. 1992). Ehhez hasonlóak az óvodákban használatos mérési és értékelési programcsomagok, amelyek a gyermekek nyelvi kompetenciáját, verbális képességeit és azok fejlődését követik nyomon (pl. Bakonyi 2009). A gyermek anyanyelvi szintje célszerűen vizsgálható különböző tesztelési eljárásokkal is (vö. Imre 2007). A tesztek a nyelvelsajátítás több területét érinthetik: a beszédhangok artikulációját (pl. Arizona Articulation Proficiency, Fudala 2000), a szókincs gazdagságát (pl. LAPP-teszt, Lőrincz et al. 1995), vagy a grammatikai fejlettséget (pl. Nyelvfejlődési szűrővizsgálat, Pléh et al. 2002) stb. Természetes, tipikus kommunikációs szituációban rögzített spontán beszéd alapján az anyanyelvi teljesítmény különféle mérésekkel írható le. Gyakran alkalmazott mutató a nyelvelsajátítási szint megállapításában a Mean Length of Utterance (MLU), vagyis a megnyilatkozások átlagos hosszának mérője, ami megadha-

tó a morféma (MLU_m) vagy a szavak számával (MLU_w) (l. pl. Parker–Brorson 2005; Mabel et al. 2010); a Developmental Sentence Scoring (DSS) eljárással kinyert pontszám (Lee–Canter 1971), illetve magyar adaptációja, a közlésegyeségek fejlődési mutatója (KFM) (Gerebenné Várbíró et al. 1992), vagy a Language Assessment, Remediation and Screening Procedure (LARSP) által készített profil (Ball et al. 2012), és az elkészült magyar változat (Bunta et al. 2015). A kutatók számára a természetes mintavétel nem mindig egyszerű, hiszen nehéz a gyermek otthonában, ismerős környezetében megteremteni a jó minőségű rögzítés feltételeit, illetve szokatlan, eltérő környezetben kialakítani egy, a gyermek számára kellemes beszédhelyzetet.

A rögzített spontán beszéd elemzése lehetőséget adhat a gyermeknyelv változatos nyelvi elemeinek tanulmányozására. Vitatott metodológiai kérdés azonban, hogy mennyi és milyen minta szükséges a kvantitatív elemzésekhez (Tomasello–Stahl 2004). A korpusz méretét, vagyis hogy hány gyermektől származzon az adat, illetve milyen hosszú hangminta szükséges a vizsgálat-hoz, a kutatási kérdéshez igazítva kell kialakítani, annak figyelembevételével, hogy milyen a vizsgálandó jelenség gyakorisága a hétköznapi nyelvhasználatban. Az anyaggyűjtés menetének támogatására a kutatók egyre inkább töreksenek adatbázisok létrehozására. Az 1984-ben megalapított CHILDES (Child Language Data Exchange System = Gyermeknyelvi adatsere rendszer, l. MacWhinney–Snow 1985; MacWhinney 2000) a tágabb kutatóközösség számára teszi elérhetővé a különféle anyanyelvű gyermekek anyagát az interneten (<http://childes.psy.cmu.edu>). A böngészhető adatbázisban megtalálhatók többek között angol, német, dán, svéd, norvég, francia, olasz, görög és magyar gyermekektől származó átírt hanganyagok, illetőleg lehetőség van további anyagok feltöltésére is. A CHILDES anyagán készült kutatásokból eddig több mint 3000 publikáció készült (lásd <http://talkbank.org/info/usage/childesbib.pdf>).

A nemzetközi beszédkutatások számára több, főként angol és amerikai gyermeknyelvi beszédkorpusz áll rendelkezésre, amelynek jelentős része beszédtechnológiai indíttatású (beszédfelismerők számára készült). A PF STAR gyermeknyelvi korpusz (Batliner et al. 2005) egy többnyelvű (brit angol, német, olasz, svéd) adatbázis, amely több mint 60 órnyi beszédet tartalmaz, anyanyelvi és nem anyanyelvi, olvasott és spontán felvételeket. A 611 gyermek életkora 4 és 14 év között szórt. A LISTEN (Literacy Innovation that Speech Technology ENables) Reading Tutor projekt (Banerjee et al. 2003) célja, hogy automatikus beszédfelismerő segítségével monitorozzák a gyermekek hangos olvasását, valamint segítsék őket a hibák javításában hangzó és grafikus visszacsatolásokkal. Ennek a fejlesztésnek a keretein belül készült a KIDS nevű amerikai korpusz (Eskenazi 1996), amely különböző életkorú (1–3. osztályos) és olvasási képességű gyermek szövegeit tartalmazza.

Az OGI Kids’ Speech Corpus (Shobaki et al. 2000) 1100 gyermek irányított és spontán beszédét tartalmazza. A gyermekeket óvodáskortól 10. osztá-

lyos korig tesztelték, az anyag izolált szavakat, mondatokat is tartalmaz. A TBALL (Technology Based Assessment of Language and Literacy) elnevezésű project (Kazemzadeh et al. 2005) 256 olvasni tanuló, angolul beszélő gyermekek anyagát tartalmazza. Az adatközlők 5 és 8 év közöttiek, 69%-uk spanyol anyanyelvű, 24%-uk angol anyanyelvű, 5%-uk angol-spanyol kétanyelvű beszélő. A 30 000 felvétel összesen 40 órát tesz ki. Az egy gyermektől felvett maximum húsz perces anyagok között megtalálható az olvasás, a szín- és a képmegnevezés is. Feljegyezték emellett a gyermekek életkorát, évfolyamát, angol nyelvi szintjüket, anyanyelvüket, otthon és barátokkal használt nyelvüket és a szülők anyanyelvét is. A hanganyaghoz fonetikai transzkripció is készült.

Magyar nyelvre az utóbbi években indult meg a gyermeknyelvi korpuszok építése. Noha 10–16 év közötti falusi gyermekektől származó hangfelvételeket (meséket, elbeszéléseket) már a Hegedüs-archívumban is találunk, amelyek az 1940-es évekből származnak (Menyhárt 2012). A SPECO projekt keretein belül, beszédfeldolgozási alkalmazások számára készült adatbázis 5–10 éves gyermekek által bemondott (ismételt, illetve olvasott) hangkapcsolatokat, szavakat, mondatokat tartalmaz (Csatári et al. 1999). A Magyar Óvodai Nyelvi Korpusz (MONYK) 4,5–5,5 éves, különböző szociokulturális háttérrel rendelkező budapesti gyermekek 20–30 perc hosszúságú felvételeit tartalmazza (Mátyus–Orosz 2014). Az elkészült interjúkhoz a CHILDES szabályai szerinti átírás, valamint morfoszintaktikai annotálás is készült. A jelenleg fejlesztés alatt álló GABI (Gyermekegyesítő AdatBázis és Információ-tár) egy széles életkori spektrumot átfogó, sok szempontú kutatásra alkalmas, nagy mennyiségű hanganyagot tartalmazó gyermekbeszéd-adatbázis (Bóna et al. 2014). Ez az adatbázis 3–18 éves gyermekek/fiatalok 30–40 perces beszédfelvételeit rögzíti. A fejlesztők a protokoll kidolgozásához a BEA adatbázis (Gósy et al. 2012) protokollját vették alapul. A fejlesztés 2013 februárjában kezdődött, azóta több mint 200 felvétel készült el a gyermekek természetes környezetében, az óvoda, iskola, esetleg saját otthon egy csendes termében (egy gyermektől átlagosan korcsoporttól függően 25–35 perc).

Minthogy a különböző nyelvek beszédatadabázisai elsősorban a kisgyermekek, illetve a felnőttek beszédére fókuszálnak, felmerült az igény, hogy a két csoport közé eső életkori szakasz is reprezentálva legyen, ezáltal beszédük megfelelő módon kutathatóvá váljék. A jelen tanulmány egy fejlesztés alatt álló, magyar nyelven egyedülálló adatbázis kezdeti munkálatait, tartalmát, felvételi körülményeit és lejegyzési stratégiáit ismerteti. A Tini BEA adatbázis, mint ahogy a neve is mutatja, a BEA adatbázis kiegészítő adatbázisa, amely a felnőtt BEA és a gyermeknyelvi adatbázisok közötti hiányt hivatott lefedni.

A Tini BEA elődje: a BEA adatbázis

A BEA Spontán Beszéd Adatbázis fejlesztése 2007 őszétől folyamatosan zajlik a Magyar Nyelvtudományi Intézet Fonetikai Osztályán. A többfunkciós

beszélt nyelvi korpusz létrehozását mindenekelőtt a fonetikai elemzések igénye, illetve a beszéd kutatás új feladatai indokolták; ezek ugyanis elképzelhetetlenek lennének egy megfelelő hangminőségben rögzített, finom fonetikai elemzésekre is alkalmas adatbázis nélkül. Az alapkutatások támasztotta elvárásoknak köszönhetően a BEA adatbázis (nevének némiképp ellentmondva) nem csupán többféle (narratíva, véleménykifejtés, tartalomösszegzés, háromfős társalgás) spontán beszédet tartalmaz, de megtalálhatóak benne mondat- és szövegfelolvasások, illetve mondatismétlések is. A korpusz adatközlői 20 évnél idősebb, egynyelvű, budapesti, köznyelvet beszélő felnőttek; különböző iskolázottsági fokkal (nyolc általános, középiskola, felsőfokú végzettség) és foglalkozással (pl. tanár, színész, lelkész, autószerelő, mérnök, bankár, orvos, pókerjátékos stb.). Az adatbázis nagysága napjainkra meghaladta a 350 beszélőt, ami mintegy 300 órnyi hanganyagot jelent. A korpusz legfőbb előnyei közé tartozik, hogy időt takarít meg a kutatók számára az által, hogy nem nekik kell felkeresni az egyes adatközlőket, és elkészíteni a megfelelő és ugyanolyan minőségű felvételeket (ami felvételi helyiség és készülék hiányában sokszor lehetetlen vállalkozás lenne), illetve az adatbázis nagyságának köszönhetően a beszélők száma sem korlátozott többé.

A különböző beszédkorpuszok nem csupán a hangzó anyagot, de általában annak átíratát is tartalmazzák, így a BEA nagy mennyiségű hangzó anyagának lejegyzése a korpusz fejlesztésének megkezdése óta folyamatos. A felvételek átírata a felhasználási területtől függően lehet helyesíráson alapuló lejegyzés, fonetikai transzkripció, tartalmazhatja az intonáció és egyéb szuprasegmentumok jelölését. A BEA hanganyagainak lejegyzése kezdetben egy elsődleges írásos tükröztetés volt. Ezek az elsődleges átíratok a Microsoft Office Word programjában .doc formátumban, helyesírásban, központosítás nélkül készültek, a későbbi feldolgozás szempontjából fontosnak ítélt adatok, mint például a megakadásjelenségek, illetve a fiziológiai hangadások jelölésével. A helyesíráson alapuló lejegyzés nem jelölte a kiejtés és a helyesírás eltéréseit, tehát nem érvényesítette a hasonulási, összeolvadási szabályokat (a *szabadság* szó nem *szabaccság*ként lett lejegyezve). A megakadásjelenségek jelölése a vastagon szedett alak volt, és ha nem hangzott el javítás, a helyes szóalak []-ben lett megadva, például: *kell még tenyér meg kej [kenyér meg tej]*. A lejegyzésben jelölve voltak a néma és kitöltött szünetek, nyújtások, a spontán beszédben gyakran használt, nem szótári alakban előforduló szavak (pl. *asszem, nem tom*), az idegen szavak, rövidítések, betűszók, mozaikszók, illetőleg a lejegyző számára értelmezhetetlen szóalakok. Az elsődleges lejegyzések a kutatók munkáját megkönnyíteni hivatott durva átíratok voltak. A későbbiekben a kutatási igények, illetve az automatikus, gépi beszéd felismerésben való felhasználás szükségessé tették a lejegyzési elvek újragondolását, és a szoftveres háttér megváltoztatását. 2010 októberétől a Word dokumentumban történő átírást felváltotta a Transcriber szoftverrel történő lejegyzés, és ezzel egyidejűleg megkezdődött a már elkészült Wordben lejegyz-

zett anyagok Transcriberbe való átkonvertálása. A Transcriber programban elkészült lejegyzések nagy előnye, hogy az írott szöveg és a hanganyag egyszerre látható és hallható; a program lehetővé teszi a beszéd szegmentálását, címkézését és leírását (Barras et al. 1998). (A BEA transcriberes lejegyzéséről bővebben l. Gyarmathy–Neuberger 2011.)

Noha az imént röviden bemutatott lejegyzési mód az adatbázis beszédfelismeréshez történő felhasználásához kiválóan megfelel, a kutatásban való felhasználhatósága korlátozott. Nem alkalmas ugyanis fonetikai mérésekre, elemzésekre. Ahhoz, hogy az adatbázis eredeti céljainak megfelelően minél szélesebb körben felhasználhatóvá váljék, kidolgozásra került egy újabb, a Praat szoftverrel (Boersma–Weenink 2013) történő lejegyzési rendszer. Ez a típusú transzkripció három szinten történik: beszédszakasz, szó- és hangszinten, így minden beszélőhöz 3-3 lejegyzési sáv tartozik az átiratban. A beszédszakasz szintű átirat helyesírásban, központosítás nélkül történik, a szószintű tartalmazza a kiejtett szavak szótári alakját, a hangszintű pedig a ténylegesen elhangzott beszédhangokat. A tipikustól eltérő beszédhang-realizációk (például irreguláris zöngé vagy hehezetes hangok) speciális karakterrel vannak jelölve. A korábbi lejegyzésekhez hasonlóan itt is jelölve vannak a megakadásjelenségek, az átfedő beszéd, a különféle zajok, hűmmögés, nevetések stb. Ez a fajta átirat lehetőséget ad a fonetika, a pszicholingvisztika és a beszédtechnológia területén belül olyan szempontok vizsgálatára, mint a beszéd szegmentális és szuprasegmentális jellemzői, a megakadásjelenségek, a beszédet kísérő nem verbális elemek, bizonyos jelenségek automatikus osztályozása vagy az átfedő beszéd detektálása.

A Tini BEA fejlesztésének kezdetei

Ahogy azt a bevezetőben is említettük, az elmúlt évtizedekben elsősorban gyermek- és felnőtt nyelvi korpuszok születtek. A köztes időszakot, a tinédzserek nyelvhasználatát rögzítő adatbázis azonban (magyar nyelven) eddig nem létezett. A tizenévesek beszédének vizsgálata, az erre az életkori csoportra jellemző finom változásokat feltáró, illetve a tinédzserek beszédét a felnőttekétől elkülönítő tényezők megismerésére irányuló kutatások szükségessé tették egy a BEA adatbázishoz hasonlóan a modern hangrögzítés minden kritériumát teljesítő, fonetikai, alkalmazott fonetikai és pszicholingvisztikai kutatásokra egyaránt alkalmas adatbázis létrehozását. Az ún. Tini BEA adatbázis a BEA Spontán Beszéd Adatbázis kiegészítéseként jött létre, középiskolás korú fiatalok spontán beszédét rögzíti. Az adatbázis fejlesztése 2013 utolsó harmadában indult az MTA Nyelvtudományi Intézetének Fonetikai Osztályán. A felvételek készítése során figyelembe vesszük a Nyelvtudományi Intézetben érvényes *Humán vizsgálatokon alapuló nyelvészeti kutatások etikai szabályozásában* foglaltakat (2007. április 16 – Az „1995. évi CXIX. törvény a kutatás és a közvetlen üzletszerzés célját szolgáló név- és lakcímadatok kezeléséről”, valamint az „1992. évi LXII. törvény a személyes

adatok védelméről és a közérdekű adatok nyilvánosságáról” figyelembevételével. A hivatkozott törvények a Nyelvtudományi Intézet belső honlapján is hozzáférhetők). Az adatközlők anonimizálva, kódokkal szerepelnek az adatbázisban. A kódok megmutatják a felvétel sorszámát, és az adatközlő nemét, de személyük azonosíthatatlan. A korpusz adatközlői 16–18 éves budapesti gimnazisták; a későbbiekben azonban tervezzük a más iskolatípusokban tanuló diákokra való kiterjesztését. A korpuszban mind a nemek arányát, mind az életkorokat igyekszünk kiegyenlíteni.

A Tini BEA protokollja

Az adatbázis hangfelvételei meghatározott protokoll szerint készülnek, amely csaknem teljes mértékben megegyezik a BEA adatbázis felvételi protokolljával. Ez a rögzítendő hanganyag részeit, azok tartalmi sajátosságait, illetőleg az egyes részekben részt vevő személyek számát tartalmazza. A különféle típusú spontán beszédnél ebben a korpuszban is helyet kaptak a mondatismétlések, valamint a mondat- és szövegfelolvasások. A beszélők életkori sajátosságait, illetve érdeklődését figyelembe véve egyetlen felvételi részen változtattunk: a Tini BEA adatbázisban a társalgás helyett egy ún. „zsebszöveg” szituációs játék kapott helyet (a játék részletes leírását lásd az 5. pontban). Mivel a zsebszövegben nem a felvételvezető és egy diák, hanem két diák vesz részt, az adatközlők a BEA adatbázissal ellentétben párosával érkeznek a felvételre. További különbség a két adatbázis között, hogy a Tini BEA-ban (épp a zsebszövegből fakadóan) az egy beszélgetésben részt vevő személyek száma maximálisan két fő. Ennek a szerkezeti átalakításnak köszönhetően két felvételhez tartozik egy zsebszöveg rész. A hat részből álló protokoll a következő: mondatismétlés, narratíva, véleménykifejtés, tartalomösszegzés, zsebszöveg, felolvasás.

1. A mondatismétlés 25 változatos grammatikai szerkezetű, szórendű, és koartikulációs szabályokat tartalmazó mondatot foglal magában. A tesztmondatok 3-4 szótagú szavakat tartalmaznak, átlagosan 8-12 darabot.

2. A narratíva során a diákok életükről, iskolájukról, iskolán kívüli tevékenységeikről, hobbijaikról mesélnek; ezek általában összefüggő, monologikus szövegek.

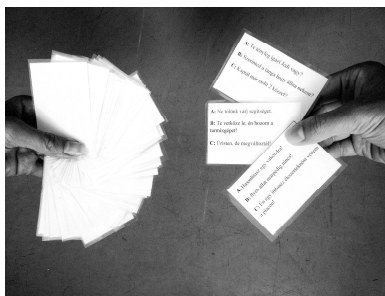
3. A véleménykifejtés szintén többé-kevésbé monologikus jellegű. A felvételvezető ebben a részben különösen ügyel arra, hogy kifejezetten a kamaszok érdeklődési körébe tartozó dolgokról kérje ki a véleményüket, az egyes témák a diákok mindennapi életét érintik. A témakörök nemenként is különbözhetnek. A lányok esetében például olyan kérdéskörök fordulnak elő, hogy jónak tartják-e, ha valaki sminkeli magát az iskolába, náluk a házirend kitér-e erre, a tanárok miként viszonyulnak ehhez, ők mit tartanak helyesnek; vagy hogyan illik öltözködni az iskolába, ehhez képest ők miket tapasztalnak, vannak-e kirívó esetek. Az inkább a fiúkat érintő témák például, hogy mennyire befolyásolja az osztálybeli népszerűséget, hogy kinek milyen típusú mobilte-

lefonja van, mennyire márkás ruhákat hord, számít-e a társak közti megítélésben a szülők anyagi helyzete; vagy mit gondolnak az okostelefon ill. mobiltelefon használatáról az iskolában, akár óra közben, szabályozva van-e ez náluk, vannak-e erre vonatkozó szankciók. Vannak mindkét nemet érintő, univerzális témák, mint például: mennyire törekszik egészséges dolgokat árulni az iskolai büfé, ők mennyire igénylik ezt, kellene-e valamit változtatni; vagy fontosnak tartják-e a pár éve bevezetett mindennapi testnevelést, mennyire valósul meg az iskolájukban; továbbá mit gondolnak a külföldön való továbbtanulásról. Ebben a részben az előzőhöz képest némiképp már nehezebb a tanuló feladata, hiszen át kell gondolnia, mi a személyes állásfoglalása az adott dologgal kapcsolatban, ebből mennyit kíván nyilvánosan felvállalni, mi az általános vélekedés, mit illik gondolni az adott témáról, stb. A vélemény megfogalmazása azonban nem igényel háttértudást, konkrét ismereteket. A minél hosszabb megnyilatkozást elősegítendő előfordul, hogy az interjúkészítő provokatív célzattal közbeszól, egy az adatközlőével ellentétes, vagy azzal azonos véleményt (attól függően, hogy melyik készíti a tanulót hosszabb megnyilatkozásra) képviselve.

4. A tartalomösszegzés rész irányított spontán beszédet tartalmaz; az adatközlő két szöveget hallgat meg felvételtől, amelyeket ezután saját szavaival kell elmondania. A két szöveg, egy rövid tudományszerűsítő cikk, illetve egy történelmi anekdota, megegyezik a BEA esetében használttal. Ez a rész a diákok számára talán kevésbé nehéz, mint a felnőtteknek, hiszen az ő mindennapi iskolai életükben (pl. nyelvórákon) gyakran találkozhatnak olyan feladatokkal, amikor egy elhangzott szöveget össze kell foglalniuk.

5. A zsebszöveg rész egy nagyfokú kreativitást és együttműködést igénylő szituációs játék, amelyben két tanuló vesz részt. A felvételvezető csak min-tegy néma játékvezető van jelen, a szituáció alakulását, megvalósítását nem befolyásolja. Ennek a játéknak a lényege, hogy a két adatközlő egy kártyapakliból húz 3 kártyát, amin különféle mondatok olvashatóak. Minden kártyán három (A, B, C) mondat szerepel. A felvételvezető előre meghatározza, hogy melyik betűjelű mondatot kell felolvasniuk a tanulóknak. A mondatokat nem szabad előre megnézniük, hiszen a játék az oda nem illő megnyilatkozások okozta humoros helyzetekre épül. A felvételvezető előre felvázol egy (a kamaszok hétköznapijaiba illeszkedő) szituációt, amit ők elkezdenek előadni (pl. egy diák iskolába menet látszólag magában beszél, amikor hirtelen rákös-szön az osztály legnépszerűbb tanulója). Amikor már a megadott szituáció kibontakozott, a felvételvezető jelet ad (rámutatással), és az adott szereplő felolvassa a nála lévő kártyán szereplő mondatot. A játék lényege, hogy a játékosoknak ezt a mondatot úgy kell kezelniük, mintha a beszélgetésben tényleg felmerült volna. A kártyán álló mondatok a következők lehetnek: *Milyen így most, hogy átoperáltak?* vagy *Kössetek az árbochoz, és ne adjatok rumot!* vagy *Azt hiszem, lenyeltem egy denevért.* stb. (1. ábra). Összesen több mint

százötven mondat kerülhet a játékba, így a szituáció kimenete minden felvételi esetében eltérő.



1. ábra

A Tini BEA „zsebszöveg” szituációs játékának kellékei

6. Az utolsó részben az adatközlők feladata felolvasás. Elsőként azt a 25 tesztmondatot kell felolvasniuk, amit az első részben megismételtek, majd egy 234 szóból álló tudománynpszerűsítő szöveget. Ez a rész a tapasztalatok szerint a diákoknak szintén kevesebb nehézséget okoz, mint a felnőtteknek, hiszen nekik még a napi órai rutinhoz tartozik a hangos olvasás.

A felvételi körülmények

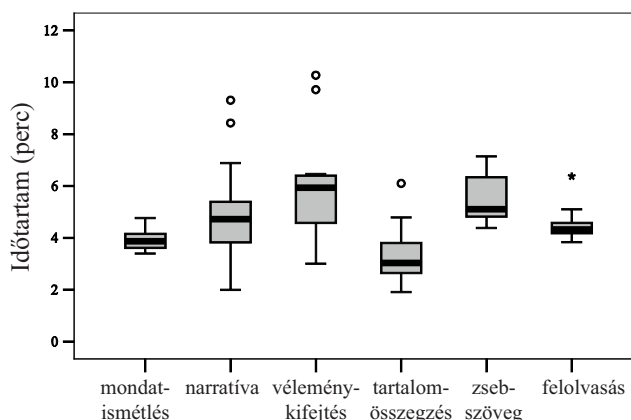
A legfontosabb követelmény a Tini BEA adatbázis esetében is a megfelelő jel-zaj viszonyú, széles frekvenciatartományú, magas dinamikájú, torzításmentes felvételek elkészítésének biztosítása volt. Így ennek a korpusznak a hangfelvételei is az MTA Nyelvtudományi Intézet Fonetikai Osztályán található zajszigetelt szobájában készülnek. A kritériumainknak megfelelő technikai háttérrel az Audio-technika AT4040 típusú kardiod kondenzátor mikrofonok, a Phonic MM102 típusú 2 csatornás phantom tápos analóg keverőpult, továbbá a GoldWave hangeditáló szoftverrel, 44,1 kHz-es mintavételezéssel közvetlenül a számítógépre történő digitális rögzítés biztosítja.

Leíró statisztikai adatok

Adatközlőink egynyelvű, köznyelvet beszélő budapesti középiskolások, 10 fiú, és 8 lány, átlagéletkoruk 16,44 év (16–17 év). A beszélők két budapesti gimnáziumból érkeztek, de a közeljövőben tervezzük az adatbázis további iskolatípusból érkező adatközlőkkel való bővítését is.

A Tini BEA adatbázis jelenleg 18 felvételt tartalmaz, ami 7 óra, 34 perc 30 másodpercnyi hanganyagot jelent. A legrövidebb felvétel 21 perc és 36 másodperc hosszú, míg a leghosszabb 39 perc 40 másodperc. Egy felvétel átlagosan 28 perc 1 másodperc időtartamú.

Az egyes felvételrészek különböző időtartamúak (2. ábra). Átlagosan a leghosszabb (5 perc 47 másodperc) a véleménykifejtés rész, amit a zsebszöveg követ 5 perc 33 másodperccel; a diákok ebben a két részben bizonyulnak a legbeszédesebbnek. Valamivel rövidebb átlagos időtartammal (5 perc 1 másodperc) valósul meg a narratíva; a mondatismétlés (átlag: 3 perc 56 másodperc) és a felolvasás (átlag: 4 perc 28 másodperc) időtartamát pedig nagyrészt a megismétlendő, illetőleg felolvasandó szövegek hossza határozza meg. A legrövidebb rész (3 perc 15 másodperc) átlagosan a tartalomösszegzés, ezen belül is a tudománynépszerűsítő szöveg ismertetése (1 perc 21 másodperc), ami magyarázható egyrészt a feladat komplexitásával, másrészt azzal, hogy a hallott szöveg hossza, illetve az abban elhangzott tartalmi elemek száma eleve behatárolja az elmesélés hosszúságát. A tudománynépszerűsítő szöveg hossza 1 perc 37 másodperc, míg a történelmi anekdota 2 perc 6 másodperc.



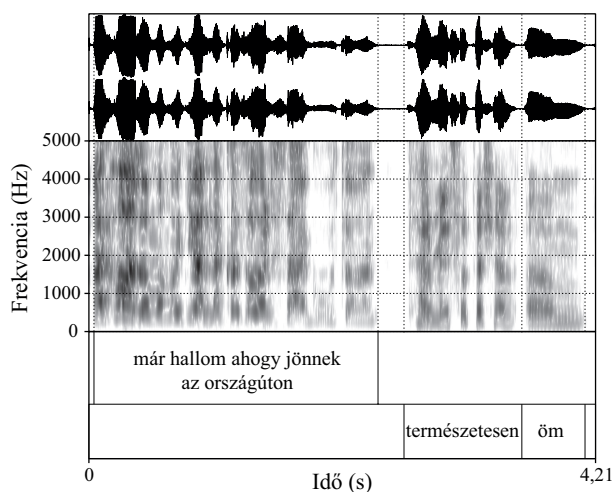
2. ábra

Az egyes felvételrészek időtartama (medián és szórás)

Az egyes beszélők egyéni különbségei az időtartam tekintetében a legjobban a narratíva és a véleménykifejtés részekben mutatkoznak meg. A legkevesbé beszédes adatközlőnk narratívája mindössze 1 perc 59 másodpercet, véleménykifejtése 3 percet tett ki (érdekességgént megemlíjtük, hogy az általános vélekedésnek némiképp ellentmondva ő egy női beszélő). Önmagáról a leghosszabban (9 perc 18 másodperc) egy fiú adatközlő mesélt, míg a felvéltelvezető által elővezetett témáról egy lány ismertette leghosszabban (10 perc 16 másodperc) a véleményét.

A hanganyagok lejegyzése

Az első felvételek elkészülése után elkezdődött a hanganyagok lejegyzése is. Az átíratózás időillesztett, a Praat programban történik. Első lépésként a felvételek beszédszakasz szintű annotációjára terjed ki (3. ábra). Ez a lejegyzés később tovább bővíthető a BEA harmadik lejegyzési útmutatója szerinti szó- és hangszintű címkékkel (lásd korábban).



3. ábra

A Tini BEA beszédszakasz szintű lejegyzése a Praat programban

A beszédszakasz szintű címkék helyesírásban (fonéma alapon) lejegyezve tartalmazzák a szünettől szünetig terjedő szöveges részeket (beszédszakaszokat), valamint a kitöltött szüneteket (az elhangzott hangkapcsolattal jelölve, például *öm*, *öő*). Ebben az elsődleges lejegyzésben a néma szünetek nem kapnak külön jelölést, a jelkimaradások ideje alatt a címke üres marad. Az átíratban semmiféle központosítást nem alkalmazunk. Ennek fő oka, hogy a spontán beszédben a (tag)mondatnyi egységek határainak bejelölése nem lenne egységes, hiszen az a mindenkor lejegyző szubjektív értékítéletén alapul. Az egyes beszélőkhöz tartozó különálló címkesorok lehetővé teszik, hogy az egyszerre beszélések, vagyis az átfedő beszédrészek jól kivehetők legyenek, illetőleg időtartamuk is meghatározható. A beszédszakaszok címkéjén belül jelölve vannak a megakadásjelenségek (például ismétlések, újra-indítások, téves kezdések), valamint a nem verbális jelenségek (például nevetések, hűmmögések). A lejegyző számára érthetetlen vagy egyéb zaj miatt használhatatlan beszédrészek a kérdőjel (?) jelölést kapják. A köznyelvben használatos, de nem szótári alakjukban előforduló szavakat, kifejezéseket úgy

jelezzük le, ahogyan a beszélő ejti (például *asszem, nemtom*). A számokat mindig kiírjuk betűvel, a helyesírásnak ellentmondóan kötőjel nélkül egybeírva jegyezzük le (például *kétezerháromszáztízben*). A kötőjel az egyetlen írásjel, amelyet alkalmazunk a lejegyzésben; ez a valamilyen okból töredékesen ejtett szavakat jelöli. Ilyenek például az újraindítások (például *hol- hol-nap*) vagy a téves kezdések (például *elgya- elpróbáltuk*). A megakadások esetében mindig a felszíni szerkezetből indulunk ki, ennek jelenségeit jelöli ez az átírás. Így például a téves kezdéskor elhangzott fragmentumot a lejegyző nem oldja fel, ez a jelenséget vizsgálni kívánó kutató feladata lesz. Mindeddig kilenc felvétel beszédszakasz szintű annotációja készült el.

A Tini BEA tudományos értéke

A nemzetközi és a magyar beszédatadatbázisok többsége felnőtt beszélők hanganyagát tartalmazza, és a nyelvészeti és beszédkutatások legnagyobb része is erre a korcsoportra fókuszál. Napjainkban az anyanyelv-elsajátítás menetének és a gyermeknyelv sajátosságainak nagy mintán történő elemzéséhez felmerült az igény a korpuszépítésre is, amelynek során különböző életkorú gyermekekkel készültek felvételek. A gyermeknyelv és a felnőttnyelv közötti stádium mindaddig kevésbé kutatott időszaknak számított, pedig a tizenévesek nyelvhasználata és beszéde feltételezhetően eltéréseket mutat mindkét korcsoporttal szemben. Néhány eddigi magyar kutatás eredményei igazolták, hogy jellegzetes különbségeket találhatunk a tinédzserek és a fiatalabb gyermekek, illetve a felnőttek beszédének összehasonlításakor: artikulációs sajátosságokban, a spontán és az olvasott beszéd szegmentális szerkezetében (Neuberger 2015), a szuprasegmentális attribútumokban (Laczkó 2009; Valent 2010), a szókincs és a szóhasználat életkori jellegzetességeiben (Laczkó 2007), a stílusváltozatokban (Stenström 1999), a beszédprodukciós folyamatok működésében, a szóelőhívási vagy olvasási nehézségekben (Horváth–Imre 2009), a megakadásjelenségekben (Laczkó 2013) vagy a konverzációs jellemzőkben (Boronkai 2008) stb. A magyar szakirodalomban mindaddig kevés kutatás foglalkozott a tizenévesek beszédével, amelynek egyik oka lehet az ilyen korú beszélőkkel rögzített, nagy mennyiségű és rendszerezett anyagok hiánya. Ezt a hiányt kívánja pótolni a most készülő Tini BEA adatbázis, amely a kutatók számára hozzáférhető és jól használható anyagot biztosít a középiskolások beszélt nyelvének több szempontú elemzéséhez. A hanganyagokon többek között tanulmányozhatók az anyanyelv-elsajátítás utolsó, eddig kevésbé vizsgált szakaszának beszédjellemzői, valamint lehetőség nyílik a finom életkori változások feltárására. Az ilyen jellegű kutatások pedagógiai hasznossága sem elhanyagolható, hiszen a tanulás-tanítás szempontjából fontos, hogy visszajelzést kapjunk a fiatalok produkciós folyamatairól. A Tini BEA adatbázis továbbá megfelelő anyagot szolgáltathat különféle beszédtechnológiai alkalmazásokhoz, például beszédfelismeréshez, beszéd-szintézishez is.

Irodalom

- Ball, Martin J. – Crystal, David – Fletcher, Paul 2012. *Assessing Grammar: The Languages of LARSP*. Multilingual Matters, Bristol.
- Bakonyi Anna 2009. *Az óvodás gyermek fejlődésének nyomon követése. A kompetencia alapú óvodai programcsomag gyermekek megfigyelésén alapuló mérési és értékelési rendszere*. Educatio Társadalmi Szolgáltató Nonprofit Kft., Budapest.
- Banerjee, Satanjeev – Beck, Joseph E. – Mostow, Jack 2003. Evaluating the effect of predicting oral reading miscues. In: *Proceedings of EUROSPEECH 2003*. http://www.scs.cmu.edu/afs/cs.cmu.edu/Web/People/listen2/pdfs/Eurospeech2003_Evaluating_predicted_miscues.pdf (A letöltés ideje: 2014. december 12.)
- Barras, Claude – Geoffrois, Edouard – Wu, Zhibiao – Liberman, Mark 1998. Transcriber: A free tool for segmenting, labeling and transcribing speech. In: *First International Conference on Language Resources and Evaluation (LREC)*. 1373–1376.
- Batliner, Anton – Blomberg, Mats – D’Arcy, Shona – Elenius, Daniel – Giuliani, Diego – Gerosa, Matteo – Hacker, Christian – Russell, Martin – Steidl, Stefan – Wong, Michael 2005. The PF_STAR children’s speech corpus. In: *Proceedings of INTERSPEECH 2004*. 2761–2764.
- Boersma, Paul – Weenink, David 2013. Praat: doing phonetics by computer. (Version 5.3) [Computer program]. <http://www.praat.org> (A letöltés ideje: 2013. október 10.)
- Bóna Judit – Imre Angéla – Markó Alexandra – Váradi Viola – Gósy Mária 2014. GABI – Gyermeknyelvi Beszédatadátbázis és Információtár. *Beszédkutatás 2014*. 246–251.
- Boronkai Dóra 2008. Konverzációelemzés és anyanyelvtanítás I. *Anyanyelv-pedagógia* 2008. 2. szám. <http://www.anyanyelv-pedagogia.hu/cikkek.php?id=60> (A letöltés ideje: 2014. szeptember 18.)
- Bunta, Ferenc – Gósy, Mária – Bóna, Judit 2015. H-LARSP: Assessing children’s language skills in Hungarian. In Ball, Martin – Crystal, David – Fletcher, Paul (eds.): *Profiling grammar: More languages of LARSP*. Multilingual Matters. (megjelenés alatt)
- Csatári, Ferenc – Bakcsi, Zsolt – Vicsi, Klára 1999. A Hungarian child database for speech processing applications. In: *ESCA Proceedings*. BME, Budapest. 2231–2234.
- Eskenazi, Maxine S. 1996. KIDS: a database of children’s speech. *The Journal of the Acoustical Society of America* 100/4. 2759–2759.
- Fudala, Janet Barker 2000. *Arizona Articulation Proficiency Scale*. Third edition. Western Psychological Services, Los Angeles.
- Gerebenné Várbiro Katalin – Gósy Mária – Laczkó Mária 1992. *Spontán beszédmegnyilvánulások szintaktikai elemzése DSS technika segítségével*. Kézirat.
- Gósy Mária – Gyarmathy Dorottya – Horváth Viktória – Grácsi Tekla Etelka – Beke András – Neuberger Tilda – Nikléczy Péter 2012. BEA: Beszélt nyelvi adatbázis. In Gósy Mária (szerk.): *Beszéd, adatbázis, kutatások*. Akadémiai Kiadó, Budapest. 9–24.
- Gyarmathy Dorottya – Neuberger Tilda 2011. A BEA adatbázis alkalmazásfüggő lejegyzései. *Beszédkutatás 2011*. 109–120.
- Horváth Viktória – Imre Angéla 2009. A diszlexia tünetei a spontán beszédben. *Beszédkutatás 2009*. 240–55.

- Imre Angéla 2007. Az anyanyelv-elsajátítás vizsgálata. In Gósy Mária (szerk.): *Beszédészlelési és beszédmegértési zavarok az anyanyelv-elsajátításban*. NIKOL Kiadó, Budapest. 58–69.
- Kazemzadeh, Abe – You, Hong – Iseli, Markus – Jones, Barbara – Cui, Xiaodong – Heritage, Margaret – Price, Patti – Anderson, Elaine – Narayanan, Shrikanth – Alwan, Abeer 2005. TBALL data collection: The making of a young children's speech corpus. In: *Proceedings of INTERSPEECH 2005*. 1581–1584.
- Laczkó Mária 2007. Napjaink tizenéveseinek beszéde szóhasználati jellemzők alapján. *Magyar Nyelvőr* 131/2. 173–184.
- Laczkó Mária 2009. Középiskolai tanulók spontán beszédének temporális jellegzetességei. *Magyar Nyelvőr* 133. 447–67.
- Laczkó Mária 2013. A kitöltött szünetek formái és funkciója tizenévesek spontán beszédében. *Magyar Nyelvőr* 137/2. 192–208.
- Lee, Laura L. – Canter, Susan M. 1971. Developmental sentence scoring: A clinical procedure for estimating syntactic development in children's spontaneous speech. *Journal of Speech and Hearing Disorders* 36. 315–340.
- Lőrík József – Ajtony Péter – Palotás Gábor – Pléh Csaba 1995. *Az aktív szókincs vizsgálata (LAPP 3–8)*. Kézirat.
- Mabel, L. Rice – Smolik, Filip – Perpich, Denise – Thompson, Travis – Rytting, Nathan – Blossom, Megan 2010. Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *Journal of Speech, Language, and Hearing Research* 53. 333–349.
- MacWhinney, Brian 2000. *The CHILDES project: Tools for analyzing talk*. Third edition. Lawrence Erlbaum Associates, Mahwah, New Jersey.
- MacWhinney, Brian – Snow, Catherine E. 1985. The Child Language Data Exchange System. *Journal of Child Language* 12. 271–296.
- Mátyus Kinga – Orosz György 2014. MONYEEK – Morfológiailag egyértelműsített óvodai nyelvi korpusz. *Beszéd kutatás 2014*. 237–245.
- Menyhárt Krisztina 2012. A beszéd temporális jellemzői 60 évvel ezelőtti gyermek beszélőknél. *Beszéd kutatás 2012*. 246–259.
- Mervis, Carolyn B. – Mervis, Cynthia A. – Johnson, Kathy E. – Bertrand, Jacquelyn 1992. Studying early lexical development: The value of the systematic diary method. *Advances in infancy research* 7. 291–378.
- Neuberger Tilda 2015. Zöngétlen zárhangok időszerkezete az életkor és a követő magánhangzó minőségének függvényében. (megjelenés alatt)
- Ninio, Anat – Snow, Catherine E. 1996. *Pragmatic development*. Westview Press, Boulder, CO, US.
- Parker, Matthew D. – Brorson, Kent 2005. A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). *First language* 25/3. 365–376.
- Pléh Csaba – Palotás Gábor – Lőrík József 2002. *Nyelvfejlődési szűrővizsgálat (PPL)*. Akadémiai Kiadó, Budapest.
- Shobaki, Khaldoun – Hosom, John-Paul – Cole, Ronald 2000. The OGI kids' speech corpus and recognizers. In: *Proceedings of ICSLP 2000*. 564–567.
- Stenström, Anna-Brita 1999. A szlengtől a szlengbeszédig. Egy tizenévesek beszédén alapuló leírás. In Fenyvesi Anna – Kis Tamás – Várnai Judit Szilvia (szerk.): *Mi a szleng?* 89–109.

- Tomasello, Michael – Stahl, Daniel 2004. Sampling children's spontaneous speech: How much is enough? *Journal of Child Language* 31/1. 101–121.
- Vallent Brigitta 2010. Középiskolások spontán narratíváinak jellemzői. *Beszéd kutatás* 2010. 199–210.
- Vihman, Marilyn M. 1996. *Phonological development: The origins of language in the child*. Blackwell Publishing, Malden.

A kutatást az MTA Megújítás pályázat támogatta.

AZ ARTIKULÁCIÓS MOZGÁSOK AKUSZTIKAI VETÜLETEINEK ADATBÁZISA MAGYAR BESZÉDRE

Olaszy Gábor – Abari Kálmán

Bevezetés

A beszédatadabázisok egyre nagyobb szerepet kapnak a modern beszéd kutatásban és a fejlesztési munkákban is. Az adatbázisok tartalmi gazdagodása jellemzi a legújabb irányt, amikor a kutatók igyekeznek minél többféle adattal ellátni a beszédatadabázis eredeti hanganyagát. Jelen munkában is ilyen továbbfejlesztésről számolunk be. A törekvésünk az volt, hogy egy új adatbázisban számszerű adatokkal adjuk meg az artikulációs mozgások változását az akusztikai vetület függvényében mondat nagyságú jelfolyamra, annak minden pontján, minden hangjában. Így közvetve nyomon követhetők az artikulációs mozgások és állások, valamint közvetlenül vizsgálhatók azok akusztikai következményei. Akusztikai vetületnek tekintjük a zöngével gerjesztett orális csatorna rezonanciafrekvenciáit (Hz-ben megadva), amelyek az idő függvényében változva jellemzik az artikulációs mozgásokat. A beszédképzés gerjesztett szűrő modellje szerint (Fant 1960) ugyanis a beszédhangok alapvető akusztikai tulajdonságait a változó keresztmetszetű hangképző üregrendszer rezonanciátulajdonságai határozzák meg. Ezek a rezonanciafrekvenciák a legtöbb esetben a tulajdonképpeni formánsok, azonban tágabb értelmezést tesz lehetővé, ha ebben az adatbázisban továbbra is a rezonanciafrekvencia (RF) elnevezést használjuk. Ezzel biztosíthatjuk, hogy ne csak a magánhangzókra jellemző pillanatnyi rezonanciákat (formánsokat) jellemezzük, hanem le tudjuk írni például a zöngétlen mássalhangzók képzési helyét is az artikulációs állásra jellemző pillanatnyi rezonanciafrekvenciákkal. Az adatbázis sematizált formában tartalmazza a kiejtett hangsor minden hangjára

az adatokat. Egységes szerkezete lehetővé teszi programozott statisztikai vizsgálatok végzését, továbbá a koartikulációs hatások vizsgálatát, esetleges sematizálását akár három, akár öt beszédhangot felölelő területekre. Tanító adathalmazaként is használható gépi tanulási algoritmusokhoz. Tudomásunk szerint az első hasonló adatbázist, amelyik a vokális traktus rezonanciáinak adatait írja le beszédminták analizálásával angol nyelvre készítették (Deng et al. 2006).

A koncepció

Az adatbázis továbbfejlesztéséhez az alapgondolatot a gépi beszédfelismerés egyfajta általánosítása adta. Ha sok beszédből ki lehet nyerni olyan általános adatokat, amelyek alapján egy beszédjelből visszakaphatjuk az azt felépítő nyelvi egységeket (hangokat), akkor ugyanez megvalósítható a beszéd egyszerűbb reprezentálási formájával is, amikor csak a RF-adatok állnak rendelkezésre. Vagyis nem a beszédhangok azonosítására tanítjuk meg a felismerőt, hanem a RF-adatok általánosítására. Így a magyar beszédre jellemző RF-adatvonulatok valószínűsíthetők lesznek bármely hangsorra. **Ezzel elérhetjük, hogy bármely szövegre jóslhatók lesznek az RF-vonulatok anélkül, hogy azt valaki kiejtette volna.** Ennek megvalósításához tettük meg az első lépést a jelen munkával.

Anyag és módszer

A fejlesztés nyelvi alapanyaga a BME TMIT korábban fejlesztett párhuzamos, precíziós (PPBA) beszédadatbázisa (Olaszy 2013). Ebben az adatbázisban 10 beszélő ugyanazokat a mondatokat olvasta fel. A teljes adatbázis összesen 19 657 db mondatot tartalmaz. A mondatokban a hanghatárok és a magánhangzók első három formánsának precíziós meghatározása 5 mérési pontban már egy korábbi fejlesztési fázisban megtörtént, ami előfeltétele volt a jelen munkának.

A jelenlegi adatbázis kialakítása során a magánhangzókra felül a mássalhangzókkal és a jelölt szünetekkel is számolni kellett, hiszen a mondat teljes hangsorát dolgoztuk fel. A szünetjelzések közé tartozik a *start sil* és az *end sil*, amely a mondat első hangja előtti illetve az utolsó hangja utáni időintervallumot jelöli meg. A harmadik szünetjelzés a *sil*, amely a hangsor belsejében előforduló olyan események időintervallumát jelzi, amelyek nem tartoznak a fonemikus hangsorozathoz, például szünetek, idegen hangelemek. A *start sil*, *end sil* és *sil* jelzések bármelyikét a továbbiakban *SIL*-lel jelöljük. A tanulmányban a következő hangcsoportokra hivatkozunk:

- a) V = magánhangzó;
- b) $V' = V + [j \ l \ v \ m \ n \ p]$;
- c) C = mássalhangzó;
- d) $C' = C - [j \ l \ v \ m \ n \ p]$;
- e) $C'' = a \ C' \text{ szünetekkel kibővített halmaza, azaz } C'' = C' + SIL.$

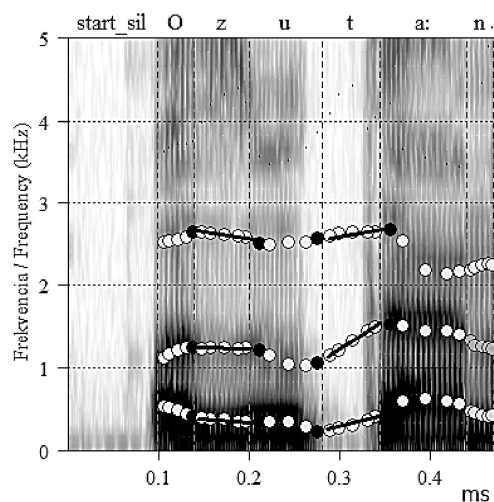
A mássalhangzók feldolgozásánál két csoportot különítettünk el. Az elsőhöz soroltuk azokat, amelyekben hasonló formánsstruktúra van, mint a magánhangzóknak (V'). Ezeket tételesen meghatároztuk az RF-adatokat. A második csoportba soroltuk a kis intenzitású, illetve a zöngétlen mássalhangzókat (C'). Ezekben csak virtuális adatokat (VRF) adtunk meg, abból kiindulva, hogy ezek az adatok a hangzásban nem vesznek részt intenzíven kis energiájuk miatt. Ezek a VRF-adatok azonban a folyamatosság és az egységes adatszerkezet szempontjából voltak fontosak. A VRF-adatok nem tükrözik az artikulációs változásokat a hangon belül.

A gépi tanuláshoz alapkövetelmény az egységes adatszerkezet. Ehhez igazodva döntöttünk úgy, hogy megtartjuk a PPBA struktúráját, azaz minden mért beszédhangra öt ponton adtunk meg adatokat, a hang 10%, 25%, 50%, 75% és 90%-os pontjában, és pontonként a 3-3 RF-adatot, de ezt kiterjesztjük a mondat minden beszédhangjára, hogy a teljes vizsgált hangsort tudjuk az orális csatorna RF-adataival jellemezni. Ez a döntés azonban új megvilágításba helyezte az RF-adatok meghatározását. A magánhangzók tekintetében természetesen az RF-adatok megfelelnek a formánsoknak (F_1, F_2, F_3). Mi legyen azokban az esetekben, amikor nem mérhetők formánsok; azért, mert például kicsi a zöngés jel amplitúdója, vagy azért, mert a gerjesztés zöngétlen? Ez utóbbi esetekre úgy határoztunk, hogy úgynevezett virtuális RF-adatokkal (VRF) fogjuk jellemezni a kérdéses jelszakaszokat. Ezek minden esetben valamely mássalhangzóhoz köthetők. A legegyszerűbb a $V'C'V'$ -kapcsolat ahol a C' -ben VRF-adatokat adunk meg. Bonyolultabb, amikor mássalhangzó-torlódás van a hangsorban ($C'C'$, $C'C'C'$, $C'C'C'C'$). A $V'_1C'V'_2$ -helyezetre vonatkoztatva azt a reprezentációs formát vezettük be, hogy a V'_1 utolsó (90%) adatait (Hz) lineárisan összekötjük a V'_2 első (10%) megfelelő adataival (Hz), és ebből határozzuk meg a hangon belüli 5 időpontban a VRF-értékeket (1. ábra).

A mássalhangzó-kapcsolódások esetére is hasonló megoldást alkalmaztunk, csak ott a $C'C'$ -kapcsolat kapcsolódási pontjára meghatározott átlagolt Hz-értékhez végeztük a lineáris összekötést (2. ábra).

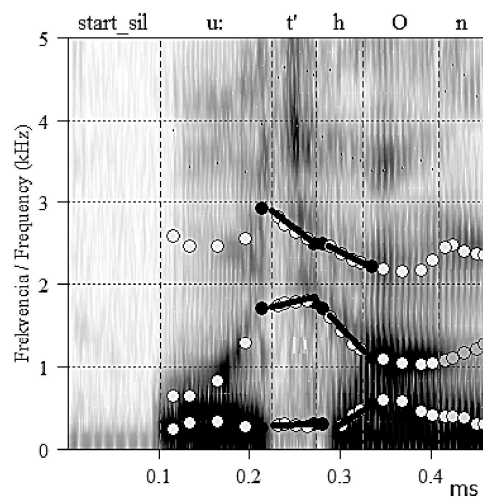
A kialakított módszerrel elértük, hogy minden hangra egységes legyen az adattér, ami áttételesen kifejezi a pillanatnyi artikulációs állásokat a beszéd folyamatában.

Az adatbázisban előforduló hangcsoportok számát és arányát az 1. táblázat tartalmazza a 10 beszélőre vonatkoztatva. Az adatbázisban tárolt hangok száma összesen: $V + C = 779\,540$ db, így az RF-ek és VRF-ek száma együttesen $11\,693\,100$ db frekvenciaérték.



1. ábra

A VRF-adatok meghatározása elméleti lineáris összekötéssel a [z] és a [t] hangokban (A kezdő és végpontokat a fekete körök jelzik. A hangokat SAMPA-jelekkel adtuk meg.)



2. ábra

A VRF-adatok meghatározása lineáris összekötéssel a [çh] C'C' kapcsolatban (A kezdő és végpontokat a fekete körök jelzik. A hangokat SAMPA-jelekkel adtuk meg.)

1. táblázat: Hangcsoportok előfordulása az adatbázisban az összes beszélőre vonatkozóan

Hangcsoport	Előfordulás (db)	Arány (%)
V	332 827	42,7
C	446 713	57,3
[j l v m n ŋ]	144 573	18,3
V'	475 400	61,0
C'	300 140	39,0

Módszer

Az RF- és VRF-adatok meghatározásához gépi és manuális kombinált módszert alkalmaztunk. Első lépésben a Praat programmal (Boersma–Weenink 2012) meghatároztuk az RF-adatokat a [j l v m n ŋ] hangokban (a magánhangzókban ezek már korábban megvoltak). Ezután vizuális ellenőrzést hajtottunk végre a rossz mérési adatok korrigálására. A tapasztalatok azt mutatják, hogy a zöngés mássalhangzók esetében az elemző algoritmusok sok hibát vétenek, ezért a vizuális ellenőrzésnek itt kiemelten fontos szerep jut. Tapasztalataink szerint a mérési hibák száma függ a beszélő hangjától is. Vannak olyan hangú bemondók, akik hangjára pontosabb értékeket állapít meg az algoritmus, és vannak olyanok, akiknél pontatlanabbakat ugyanazon beszédszakaszban. Példa látható erre a 3. ábrán.

A vizuális ellenőrzés és javítás segítségével egy célprogramot (PROFEF) készítettünk, amely lehetővé tette egyrészt az RF-adatok vizualizálását a spektrogrammal együtt a mért pontokra vonatkoztatva, másrészt a képernyőn való manuális mozgatásukat a spektrális kép és a mért RF-adatok megfigyelése alapján (Abari 2013). Egy példa látható a javítás előtti és utáni adatokra a 4. ábrán.

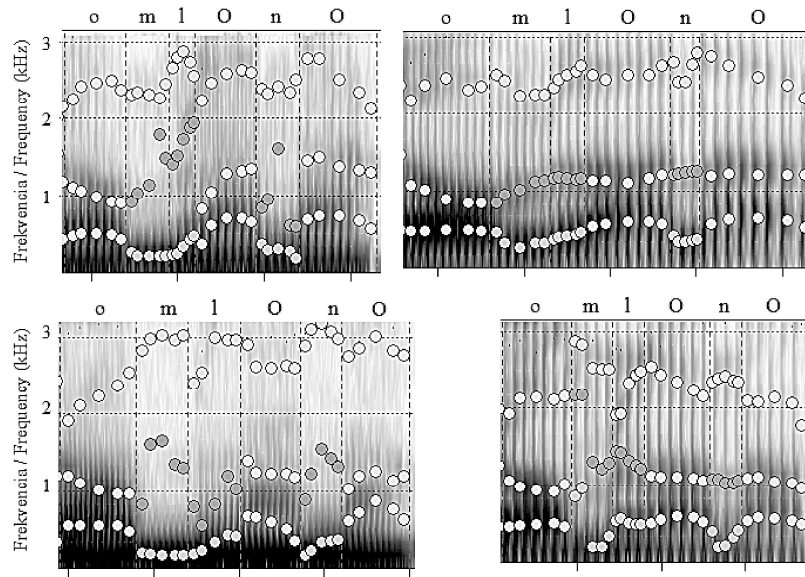
Az első lépés után korrekt RF-adatok álltak rendelkezésre a magánhangzókban és a [j l v m n ŋ] hangokban. A fejlesztés második lépését gépi segítséggel végeztük el, azaz egy erre a célra fejlesztett algoritmussal, lineáris interpoláció segítségével összekötöttük az eddig nem feldolgozott C'-hangoknál a szélső kijelölt RF-pontokat, és ebből számoltuk ki a VRF-adatokat a hang öt pontjára.

A C' virtuális rezonanciafrekvenciáit beállító algoritmus a hangkörnyezetek alapján a következő eseteket különböztette meg:

- V₁C'V₂: összekötjük V₁ 90%-os pontját a V₂ 10%-os pontjával;
- V'C'C': összekötjük V' 90%-os pontját a C' 90%-os pontjával;
- C''C'V': összekötjük C' 10%-os pontját a V' 10%-os pontjával;
- C''₁C'C''₂: összekötjük C' 10%-os pontját a C' 90%-os pontjával.

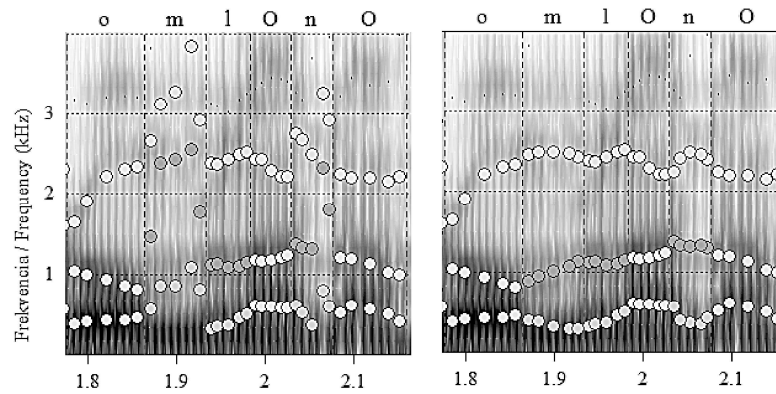
A 2. táblázatban a fenti négy eset előfordulási számát és arányát közöljük. A leggyakrabban előforduló V₁C'V₂-kapcsolat az összes eset majd 60%-át

teszi ki. A maradék kb. 40%-ban volt szükség átlagolt értékek felhasználására a VRF értékek meghatározásához.



3. ábra

A Praat programmal meghatározott RF-értékek négy beszélő hangjában a *romlana* szó [m l n] hangjaiban



4. ábra

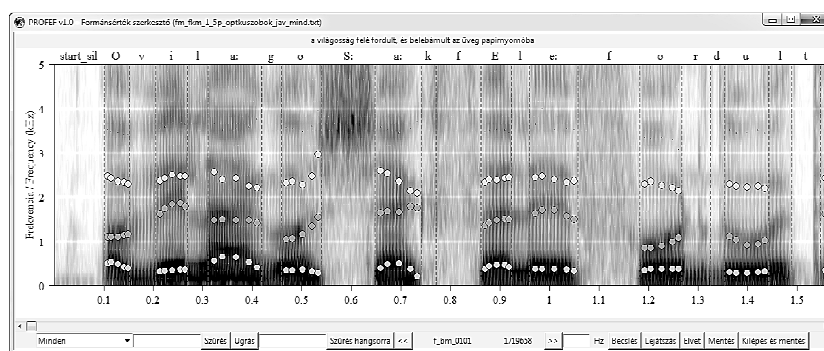
A *romlana* szó [m l n] hangjainak mért RF-értékei a javítás előtt (bal) és után (jobb)

2. táblázat: A C' VRF-adatainak beállítása során használt algoritmus négy esetének előfordulása

Esetek	Előfordulás (db)	Arány (%)
$V_1C'V_2$	177 864	58,5
$V'C'C''$	67 385	22,2
$C''C'V'$	56 015	18,4
$C''_1C'C''_2$	2 876	0,9

Az RF-, VRF-vizualizáló és -módosító PROFEF célprogram

A program vizuális grafikus felhasználói felülettel rendelkezik. Futtatásához az R statisztikai programcsomag és az Active Tcl 8.4+ ingyenes szoftvercsomag szükséges. Az 5. ábra az alkalmazás képernyőképét mutatja be, itt a képernyő legnagyobb részét a beszédmintához tartozó spektrogram teszi ki, amelyen a mért RF-értékeket kis színezett pontok jelenítik meg (a programban a RF₁ kékkel, a RF₂ pirossal, a RF₃ pedig zölddel van jelölve).



5. ábra

A PROFEF alkalmazás munkaképernyője a magánhangzók RF-adataival

A hangszínek fölött a bemondás ortografikus és fonetikus (SAMPA) formája jelenik meg. A képernyő alján a vezérlógombok és a szöveges beviteli mezők helyezkednek el. Ez utóbbiakkal meg lehet adni a keresési feltételeket. A vezérlőelemekhez tartoznak a lapozást végző gombok, amelyek segítségével a beszédadatbázis összes eleme elérhető, kereshető. Lehetőségünk van a beszédadatbázis lapozással elérhető elemeinek szűkítésére is, például egy adott beszélőre vagy egy tetszőleges hangsor, akár CVC-szekvencia előfordulására is szűrhetünk, illetve egy tetszőleges beszédmintára is ugorhatunk.

A PROFEF legfontosabb tulajdonsága, hogy segítségével vizuálisan megjeleníthető az RF-adatok sorozata (mozgása) a mondatban. A hibák felismerését a spektrogramon megjelenő, formánsértékeket jelölő pontok függőleges

pozíciójának vizuális ellenőrzésével végezhetjük el. A detektált hibák kézi javításához a pontokat az egér segítségével mozgathatjuk függőleges irányban lefelé és felfelé. A mozgítás érzékenysége 15,3 Hz/pixel, azaz a legkisebb függőleges pontelmozdulás a hangszínképen ennyivel módosítja a formánsfrekvencia értékét. A programmal akár hang, hangkapcsolat, szó keresését is elvégezhetjük. Ilyenkor a PROFEF felsorolja azokat a mondatokat, amelyekben a keresett elem szerepel, megadja, hogy hány ilyen mondat van az adatbázisban, és azt is, hogy éppen melyiket látjuk. A mondatokat egyenként megjeleníthetjük, illetve vizsgálhatjuk a kiválasztott elemet.

Eredmények

A jelenlegi adatbázis kialakítása során a V'- és C'-hangcsoportok RF- és VRF-értékeit állítottuk be. Az adatokat táblázattal tagolt szöveges állományba szerveztük, amelynek felépítése a 6. ábrán látható.

A *file* oszlopban a beszédminta azonosítója szerepel. A példában a *bm* jelű férfi (*f*) bemondó 17. könyvtárban szereplő első mondatból (*1701*) szerepel a mondat első négy hangjára [ju:li] vonatkozó adathalmaz. A *label* a hang szimbóluma SAMPA jelöléssel. A *num* azonosítja a beszédmintán belül a hangsor elemének a sorszámát (amely jelen esetben 2, mivel a mondat *start_sil* jelzéssel indul), azaz azt, hogy hányadik szegmentált elem a sorban. A *time* egy pont az időtengelyen (másodpercben mérve), amely a mérés helyét jelöli meg. A *pos* egy azonosító, amely a mért hangon belül a mérési pontokat különbözteti meg (esetünkben például a 10 a 10%-os mérési pontot jelenti). Minden hangra öt sor vonatkozik. Az utolsó három oszlop az RF₁–RF₃ értékeit tartalmazza Hz-ben kifejezve.

Az adatbázis felhasználása

Az adatbázis elsősorban gépi tanuló algoritmusok támogatását szolgálja tanító adathalmazként. Ezen felül számos más kutatás támogatására is szolgálhat. Például az ultrahanggal rögzített nyelvvalak, illetve az ebből származtatott nyelvkontúr összehasonlítható a beszédjelben mért formánsértékekkel (Csapó–Csopor 2015). Célzott keresésekkel tanulmányozhatunk adott koartikulációs mozgássorokat, elemezhetjük a koartikuláció hatáskörét. Ez utóbbira mutatunk be egy példát.

A koartikuláció hatásköre

„A koartikulációs hatás több hangot is érinthet, nemcsak a szomszédosat” – írja Gósy (2004: 149). Az adatbázis vizuális ellenőrzése során is felfigyeltünk arra, hogy bizonyos hangkapcsolatoknál a vizsgált hangot követő +2. hang is befolyást gyakorol a hang artikulációjára. Ez főleg az RF₂ adataiban látható. Magyarra ilyen vizsgálatokat még nem végeztek. Ezért célzott vizsgálatot terveztünk egy gyakori hangkapcsolatra. Az [or] kapcsolat [o] hangját vizsgáltuk, annak is a végét a 90%-os ponton, a hangkapcsolatot követő magánhangzó függvényében. Artikulációs szempontból tehát az elemzés a nyelv

vízszintes mozgására vonatkozott, amikor az előrefelé, a dentalveoláris képzési helyhez közelít. Mindezt a követő magánhangzóra (+2. hang) jellemző nyelvmozgás függvényében vizsgáltuk. Az adatgyűjtésre mutat példát a 7. ábra.

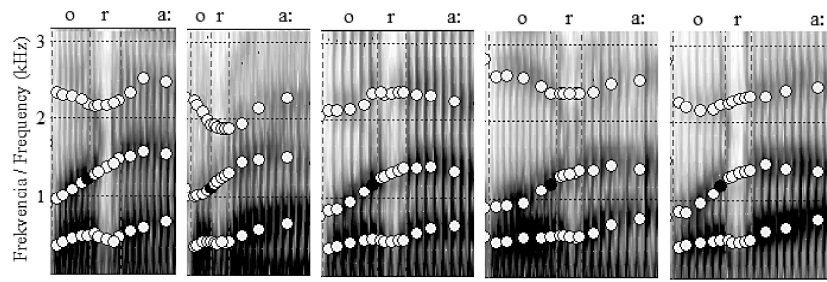
	file	label	num	time	pos	RF ₁	RF ₂	RF ₃
1	f_bm_1701	j	2	0.1067	10	263	2014	2566
2	f_bm_1701	j	2	0.1142	25	289	1998	2529
3	f_bm_1701	j	2	0.1266	50	332	1917	2492
4	f_bm_1701	j	2	0.1391	75	357	1831	2399
5	f_bm_1701	j	2	0.1466	90	366	1536	2323
6	f_bm_1701	u]	3	0.1642	10	360	1230	2241
7	f_bm_1701	u]	3	0.1831	25	328	938	2303
8	f_bm_1701	u]	3	0.2146	50	304	899	2362
9	f_bm_1701	u]	3	0.2461	75	327	1018	2404
10	f_bm_1701	u]	3	0.2650	90	361	1234	2371
11	f_bm_1701	l	4	0.2821	10	343	1425	2377
12	f_bm_1701	l	4	0.2888	25	300	1418	2368
13	f_bm_1701	l	4	0.3000	50	311	1475	2439
14	f_bm_1701	l	4	0.3112	75	301	1485	2498
15	f_bm_1701	l	4	0.3179	90	302	1479	2550
16	f_bm_1701	i	5	0.3292	10	396	1655	2548
17	f_bm_1701	i	5	0.3393	25	399	1774	2550
18	f_bm_1701	i	5	0.3562	50	388	1965	2541
19	f_bm_1701	i	5	0.3730	75	382	2107	2594
20	f_bm_1701	i	5	0.3832	90	383	2121	2617

6. ábra

A *Júliával* kezdetű mondat első hangjainak adatsora

Hipotézis: a vizsgált [or] hangkapcsolatokban az [o]-ban mérhető RF₂-értékek a hang végén annál magasabbak, minél magasabb az [r] hang után következő magánhangzó RF₂-értéke.

A vizsgálathoz az öt férfi beszélő anyagából gyűjtöttük ki a hangkapcsolatokat (3. táblázat). Összesen 114 ilyen mintát találtunk.



7. ábra

A [kora:] hangsor [o] hangja ugyanazon mondatban az öt férfi beszélő hangjában (A fekete pont a mérés helye minden mintában.)

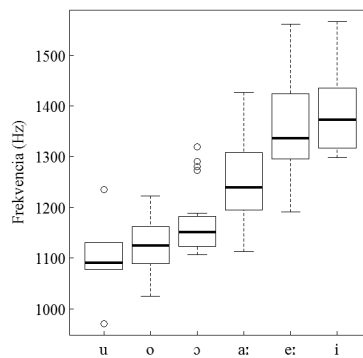
3. táblázat: A vizsgált [or] hangkapcsolatok adatai

+2. hang az [o] után	A 90%-os pont RF ₂ - átlaga az [o]-ban (Hz)	Darab	Példa
[a:]	1256	34	<i>korábban</i>
[ɔ]	1171	20	<i>életkora</i>
[o]	1129	20	<i>megkoronáz</i>
[u]	1101	5	<i>szorul</i>
[e:]	1350	15	<i>mikor ébren</i>
[i]	1390	20	<i>akkoriban</i>

A mérés eredménye szerint az [o] hangban a hang 90%-os pontján az RF₂ értéke követi a +2. hangban megjelenő magánhangzók F₂ szerinti szokásos elrendezését: u < o < ɔ < a: < e: < i (6. ábra). A 3. táblázatból kiolvasható, hogy a mért értékek átlagai rendre: 1001 Hz < 1129 Hz < 1171 Hz < 1256 Hz < 1350 Hz < 1390 Hz. A +2. hangok hat csoportjában az RF₂-értékek szignifikánsan eltérnek (Jonckheere–Terpstra-próba: $p < 0,001$). A hipotézis tehát igaznak bizonyult.

Egy további vizsgálatot végeztünk az [ɔŋk] hangkapcsolat [ɔ] hangjára az öt követő +3. hang befolyásának megállapítására az [ɔ] RF₂ 90%-os pontjának összehasonlításaival. A hipotézis ugyanaz volt, mint az előző vizsgálatban. Háromféle hangkapcsolatra találtunk adatot: [ɔŋko] (25 db), [ɔŋke] (20 db) és [ɔŋki] (5 db). Az átlagok rendre: 940 Hz, 1131 Hz és 1214 Hz. Az eltérések szignifikánsak (Jonckheere–Terpstra-próba: $p < 0,001$), tehát ez a hipotézis is beigazolódott.

A két mérésből levonható következtetés, hogy érdemes mind a +2., mind a +3. hang hatását is szisztematikusan megvizsgálni, ha a koartikuláció hatás-körét pontosabban akarjuk megismerni. Hasonló vizsgálatok elvégzését tervezzük minden magánhangzóra a jövőben.



6. ábra

Az [or] hangkapcsolat [o] hangjában, a hang 90%-os pontján az RF₂-értékek eloszlása a +2. hang (vízszintes tengely) hat csoportjában

Összefoglalás

A bemutatott adatbázisban számszerű adatokkal adjuk meg az artikulációs mozgások változását az akusztikai vetület segítségével mondat nagyságú jel-folyamra, annak minden pontján, minden hangjában tíz beszélő felolvasásos anyaga alapján. Az adatbázis adatai rezonanciafrekvenciákat jelentenek Hz-ben kifejezve. Ilyen adatbázis még nem készült a magyar beszédre. Az adathalmaz közel 800 000 beszédhang adatait tartalmazza. Nagysága jó alapot jelent akár statisztikai vizsgálatokhoz, akár tanító adathalmazként gépi tanuláshoz, akár a koartikulációs hatások pontosabb megismeréséhez. Az adatok korrektségét a kézi ellenőrzés garantálja. Az adatbázisról információ található a következő honlapon is: <http://magyarbeszed.tmit.bme.hu/artikulacio>.

Irodalom

- Abari Kálmán 2013. *A formánsmozgások statisztikai vizsgálata és modellezése a magyar magánhangzóknál*. PhD-disszertáció. DE, Debrecen.
- Boersma, Paul – Weenink, David 2012. Praat: doing phonetics by computer (version 5.3.19). <http://www.praat.org>.
- Csapó Tamás Gábor – Csopor Dávid 2015. Ultrahangos nyelvkontúrkövetés automatikusan: A mély neuronhálókra alapuló AutoTrace eljárás vizsgálata. *Beszédkutatás 2015*. 176–186.
- Deng, Li – Cui, Xiaodong – Pruvenok, Robert – Huang, Jonathan – Momen, Safiyy – Chen, Yanyi – Alwan, Abeer 2006. A database of vocal tract resonance trajectories for research in speech processing. In: *Proceedings of the ICASSP 2006*. 369–372.
- Fant, Gunnar 1960. *Acoustic theory of speech production*. Mouton, The Hague.
- Gósy Mária 2004. *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest.
- Olaszy Gábor 2013. Precíziós, párhuzamos magyar beszédadatbázis fejlesztése és szolgáltatásai. *Beszédkutatás 2013*. 261–270.

TÖBBNYELVŰ VASÚTI HANGOS UTASTÁJÉKOZTATÁS KORPUSZALAPÚ TTS MÓDSZERREL

Németh Géza – Zainkó Csaba – Bartalis Mátyás –
Olaszy Gábor

Bevezetés

A pályaudvari hangos utastájékoztatás hozzátartozik a vasúti közlekedés minőségi utaskiszolgálási követelményrendszeréhez. Mindig volt ilyen szolgáltatás a kornak megfelelő technikai szinten (Klabbers 1997). Az cikkben ismertetett fejlesztés elindítását a beszédtechnológia fejlődése tette lehetővé. Az új rendszer rugalmasabb kezelést biztosít, olcsóbb az üzemeltetése, ugyanakkor stabilabban szolgáltat jól érthető, szívesen hallgatott hangminőséget. A korábbi rendszer, amely túllépte az évtizednyi kort, gyakran már nagyon egyenetlen hangzású volt (mivel a hangfelvételek bemondásainál nem gondoskodtak már a kezdetektől fogva az egységes hangszínezet, tempó, hangsúlyozás megvalósításáról), és több hetet is igénybe vett, amíg egy bemondási módosítást elkészítettek. Mindezek egyre kényelmetlenebbé tették a rendszer fenntartását, és rontották a lejátszott hang minőségét is (csuklások, más hangszín, más tempó stb.).

A kutatás követelményrendszere

A fejlesztést sok éves kutatás előzte meg, amely előkészítette a 2013-ban kiírt követelményrendszer teljesíthetőségét. A TMIT beszédkutatói csapata már évekkel ezelőtt tanulmányozta egy ilyen bonyolult rendszer szerkezeti sajátosságait, majd megépített egy kísérleti rendszert Sárospatak állomáson 2007-ben (Zainkó–Németh 2010). Ez a rendszer már hét éve működött, amikor az általános műszaki megoldásra kiírták a pályázatot. Ebben a követelményrendszer fő elemei a következők voltak.

1. A 2014-ig működő régi rendszert megközelítő hangminőségű magyar nyelvű TTS; minimum 4.0 Mean Opinion Score (MOS) érték az alábbi skála mellett: 5 = anyanyelvű emberi bemondással összetéveszthető; 4 = anyanyelvű emberi bemondásra emlékeztető; 3 = jól érthető, de gépies; 2 = éppen érthető; 1 = érthetetlen. A MOS minimum 50 emberrel mérendő, jó akusztikai körülmények között.

2. Felügyelet és kézi korrekció nélkül is működőképes legyen a magyar nyelvű TTS.

3. A beszéd szintetizátor tetszőleges beszélő hangjára legyen elkészíthető, például a megrendelő által kért/biztosított személy hangján szólaljon meg.

4. A bemondások magyar és angol nyelvűek legyenek.

5. Az angol nyelvű bemondásokban a magyar állomásnevek magyar kiejtés szerint hangozzanak el, a más országbeliek pedig az adott helység országa hivatalos nyelvének kiejtése szerint legyenek kimondva, de ez ne zavarja az angol beszéd akusztikai egységét.

Anyag és módszer

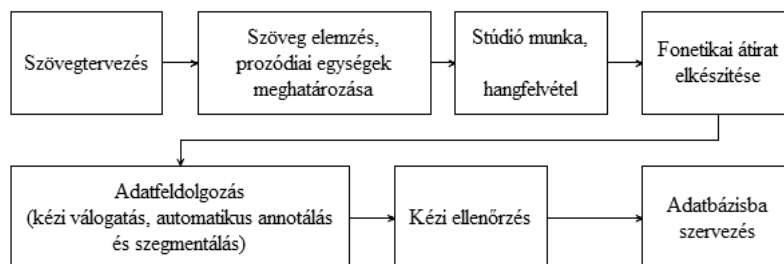
Az alkalmazott módszer a világszerte legjobb minőséget biztosító és leggyakrabban alkalmazott eljárás, a korpuszalapú hullámforma-válogatással megvalósított szöveg-beszéd átalakítás volt. Ennek kutatása már közel egy évtizede folyik a BME TMIT Beszédtechnológia és Intelligens Interakciók Laboratóriumában. A módszer lényege, hogy a felolvasási témakörnek megfelelő, azt jól lefedő, nagy mennyiségű szöveghalmazt kell meghatározni, és ezt kell felolvasatni a kiválasztott bemondóval jól kontrollált körülményeket betartva (Zainkó 2010; Yu et al. 2012). A kettő között szoros szinkront kell alkalmazni. Ez a szinkronitás biztosítja, hogy a szöveghez annak a hangzási megfelelőjét is pontosan hozzá tudjuk rendelni. A hullámformát pontos annotálással és címkézéssel kell ellátni hang- és szószinten. Ez a hullámformahalmaz képviseli a hangüzenet-beszédadatbázist, ennek kiválasztott részleteit füzi össze a szintézis során a TTS-algoritmus, amikor egy mondatot az írás alapján beszéddé alakít. A géppel összeállított hang teljes mértékben hordozza a bemondó hangkarakterét, tehát az esetek jelentős részében nem lehet eldönteni, hogy gépi hangot hallunk. A hangüzenet-beszédadatbázis elkészítése sok munkát igényel. Mind a szöveghalmaz meghatározása, mind a stúdiómunka (felolvasás) levezetése, mind pedig a hullámforma-egységek feldolgozása speciális, multidiszciplináris szakértelmet kívánt.

Hangüzenet-beszédadatbázis

A hangüzenet-beszédadatbázis képezi a hangosbemondó-rendszer adattárát. Ebből válogatva állítja elő a kívánt bemondást a rendszer. Ez az adatbázis 2998 magyar és 328 angol felolvasott mondatot tartalmaz (több fázisban növekedett, ahogy az igények bővültek). A kiválasztott bemondó a Kossuth rádió munkatársa. A stúdiómunka 110 órát tett ki, mintegy 22 órányi beszédet rögzítettünk digitalizált formában (44 kHz/16 bit lin. PCM). A rögzített hanganyag elsődleges feldolgozása gépi módszerrel (ún. forced alignment) történt, a precíz ellenőrzéseket az annotálás és címkézés tekintetében kézi erővel végeztük el minden mondaton. A felolvasandó szöveg és a hozzá tartozó hullámformák elkészítésének folyamatát az 1. ábra mutatja.

A stúdiómunka meghatározó jelentőségű egy ilyen fejlesztésnél. Itt dől el a későbbi hangzás, vagyis hogy az összefűzés után érzékelhető lesz-e, hogy a hangot részekből rakták össze vagy sem. A hangrögzítés metodológiájára saját módszert dolgoztunk ki a korábbi tapasztalatok alapján. A bemondó hangja változhat akár a napszaktól függően is, vagy a terhelés fokozásakor, illetve az egyén egészségi állapotától is függ. Biztosítani kellett, hogy az egyes

hangfelvételi alkalmakkor minden esetben ugyanazzal a stílussal, ugyanazzal a hangszínezettel beszéljen a bemondó. Gondolni kellett továbbá a sok felsorolást tartalmazó mondatok (állomásnevek) hangzásának kialakítására, ugyanakkor ezen elemek más mondatrészekkel való összefűzhetőségének rugalmas megvalósíthatóságára is. Az időpontok kimondása szintén kényes eleme egy ilyen rendszernek. Mindezekhez előzetesen hangzási kísérleteket végeztünk a bemondóval, aki bebizonyította, hogy képes akusztikailag igazodni egy korábbi hangfelvétele stílusához, hangkarakteréhez, hangszínezetéhez. A kísérletek eredménye lett a következő módszer. A felolvasandó szöveget 20 mondatnyi egységekben kezeltük, ennyit olvasott fel a bemondó egy fájlba. Az akusztikai igazodás, ráhangolódás segítésére bevezettük a mestermondat alkalmazását.



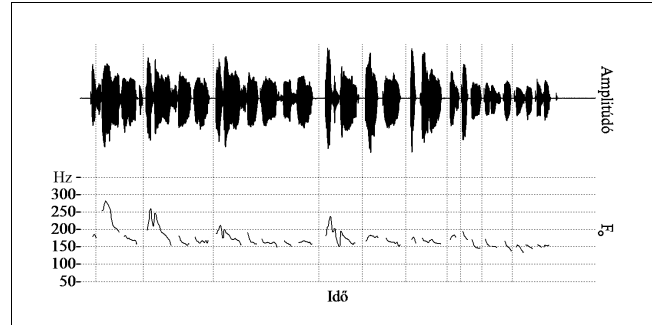
1. ábra

A hangüzenet-beszéledatbázis elkészítésének főbb lépései

Mestermondat

Az általunk bevezetett mestermondat módszert a sorozatos hangfelvételek készítése során használjuk. Célja, hogy a bemondó hangját ugyanazon szinten tartsuk, mind a beszédstílus, mind a hangszín, mind a hangfekvés tekintetében egy hosszabb hangfelvételnél, valamint további alkalmakkor is. A mestermondat szövege: *A személyvonat Abaujszántón, Abaujszántóifürdőn, Abdán, Ajkán, Akalán, és Tatán keresztül közlekedik.* A mondatot a bemondó olvasta fel. A prozódia úgy alakítottuk ki a bemondó segítségével, hogy a későbbi szintézishez optimális legyen (2. ábra). A felsorolásoknál semlegeshez közelálló kiejtést kértünk a mestermondatban, ugyanis ennek mintájára kellett a későbbiek során minden felsorolást tartalmazó mondatot felolvasnia. A mestermondatban közel lebegő dallammenettel ejtettük ki az állomás neveket minden esetben és helyzetben, első szótagi enyhe hangsúlyozással. Így biztosítottuk, hogy akár egyetlen, akár nagyon sok állomásnév is elhangozhat majd a szintetizált mondatban, mindkét esetben jól fog illeszkedni az azt megelőző és követő mondatdallamhoz. A 2. ábrán látható, hogy a mestermondat dallammenetében a felsorolt szavak végén 150 Hz-es érték körül va-

lósul meg az alapfrekvencia. Ezt a mestermondatot minden oldal elején meghallgatta és hangosan elismételte a bemondó, amíg a felvételt készítő két szakember jóvá nem hagyta a pillanatnyi hangzást, aminek ugyanolyannak kellett lenni, mint a mestermondat hangzása. Ezzel megtörtént a ráhangolódás. A jóváhagyás után indult az oldal felolvasása. Volt rá példa, hogy az oldal felolvasásának a közepén is meghallgattuk a bemondóval ezt a mondatot, amikor éreztük, hogy a hangszínezete, a beszédstílusa kezd eltérni a megkívánttól.



2. ábra

A női bemondó felolvasásban elkészített mestermondatok akusztikai diagramjai. Hullámforma (fent) és dallammenet (lent). A szavak határát függőleges vonalak jelölik

Állomásnevek

Az állomásnevek kezeléséhez is saját stratégiát alakítottunk ki. A rendszer 2031 magyar és 732 külföldi állomásnevet képes kezelni és kimondani. A magyar állomásneveknél a ragozott formák megvalósítására is gondolni kellett (például *Baja*, *Bajára*, *Bajáról* stb.). Erre külön módszert kísérleteztünk ki, mivel az nem volt követhető, hogy az összes ragozott állomásnevet (mindösszesen 14 064-et) felolvassa a bemondó (elfárad fizikailag is és hangilag is). Az állomásnevek kimondására az előírások a következők voltak.

A magyar állomásnevek kiejtése magyarul történik mind a magyar, mind az angol bemondásokban. A külföldi állomásnevek kiejtése a magyar felolvasásban az ország nyelvén történik abban az esetben, ha nincs magyar megfelelője (például *Lübeck*, *Villach*, *Würzburg*, *Mainz*). Ha van magyar megfelelője, akkor azt kell felolvasni (például *Pozsonyba*). Az angol felolvasásban a magyar állomásnevek magyar kiejtés szerint hangzanak, a külföldiek a megfelelő ország nyelvén (*The train leaves to Bratislava*).

Az állomásnevek hangfelvételéhez azt a megoldást választottuk, hogy mindegyiket egyszer felolvastattuk egy vívőmondatban, és ezt a hangfelvételt külön blokkban tároltuk, hogy a későbbiekben egyértelmű legyen a feltalálási

helye. A vivőmondat szerkezete állandó, csak az állomásnevek változnak benne (kiemelve). A vivőmondatra egy példa: *A személyvonat **Bánréve, Bánrévei vízmű, Baté, Battonya, Békéscsaba** és Tök állomásokon áll meg.* A ragozott állomásnevek vivőmondatára példa: *A felsorolt vonatok **Bajánsenyéről, Bakonyszentlászlóról, Baktalórántházáról, Balástyáról, Balatonakali-Dörgicséről** és Tatáról indulnak minden szombaton.*

A ragozott állomásnevek akusztikai megvalósításához a vivőmondat listát úgy terveztük meg, hogy tartalmazzon mintát az összes ragozott formához, vegye figyelembe a toldalékolás előtti beszédhangot, továbbá az esetleges hangváltozást (*Baja, Bajáról; Dörgicse, Dörgicsén* stb.), illetve hangkiesést (*Hegyesalom, Hegyeshalmon*) is. Így lényegesen lecsökkent a felolvasandó lista, a 14 064 helyett csak 1420 ragozott formát kellett felolvasni. Ez a ragozási lista akusztikailag teljes mértékben lefedte a 14 064 ragozott állomásnév korrekt hangzásának a megvalósítását. A ragozási lista kialakításánál a következő toldalékok csoportjait állítottuk össze:

Hol/merre? kérdésre: <i>n, on, en, ön, án, én</i>	(<i>Baján</i>)
Honnan? kérdésre: <i>ból, ből, ról, ről, áról, éről</i>	(<i>Bajáról</i>)
Hová? kérdésre: <i>ba, be, ra, re, ára, ére</i>	(<i>Bajára</i>)
Meddig? kérdésre: <i>ig, áig, éig</i>	(<i>Bajáig</i>)
Mettől? kérdésre: <i>tól, től, ától, étől</i>	(<i>Bajától</i>)
Milyen? kérdésre: <i>i</i>	(<i>Bajai kocsik</i>)

A toldalékok megelőző betű(k)re öt csoportot alakítottunk ki:

magánhangzó (V)	(<i>Abda, Abdán, Abdáról</i> stb.)
<i>j + V</i>	(<i>Türje, Türjén, Türjére</i>)
<i>...halom</i>	(<i>Kettőshalom, Kettőshalmon, Kettőshalmi</i>)
<i>z, zs</i>	(<i>Döbrököz, Döbröközre; Dúzs, Dúzsról</i>)
<i>l, r, v</i>	(<i>Almár, Almáron; Erdőszél, Erdőszélen</i>)

A fenti két csoportosítás kombinációját végigvezettük az összes állomásnéven, és a teljes listából minden előfordulásból csak egyet hagytunk meg. Így adódott ki az 1420 ragozott állomásnevet tartalmazó szűkített lista.

A szintézis során az algoritmus kiválasztja a kimondandó állomásnév alakjának hullámformáját az állomásnév blokkból, továbbá a megfelelő toldalékokat a toldalékolt állomásnév blokkból, a kettőt az adott szabály szerint összefűzi, és így valósítja meg a ragozott állomásnév hangzását. A megoldásból következik, hogy például minden *-ra/-re* toldalékoláshoz közel ugyanazt a *-ra/-re* hullámforma elemet használjuk a szintézis során. Ez azonban a hangzásban nem okoz lényeges eltérést.

Annotálás

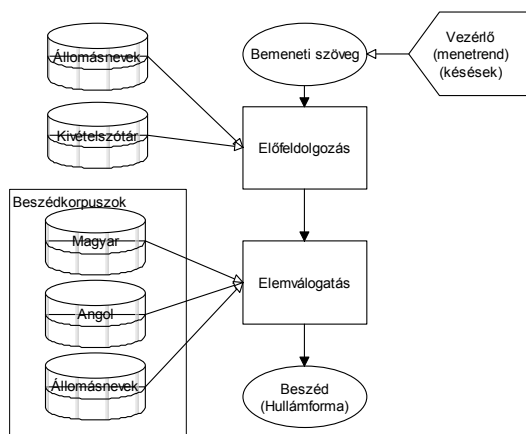
A hangfelvétel után minden (magyar és angol) mondat további feldolgozáson esett át. Elkészítettük a fonetikai átiratokat, majd gépi és kézi munkával elláttuk a mondat hangfájlját hanghatár- és szóhatárcímekkel, melyek pon-

tossága 5 ms-on belül volt a pozitívól negatívba átmenő nulla ponton. Ez azt jelenti, hogy a jelölt hanghatár a ténylegestől gyakorlatilag nem tért el. A korpuszalapú beszédszintézisnél fontos a pontos hanghatár kijelölése, hiszen az akár összefűzési pont is lehet (például a ragoknál). A szóhatár is fontos eleme a címke rendszernek, ugyanis a kidolgozott TTS-megoldásban a szintézis alapelemének a szót tekintettük. Ez azt jelenti, hogy az esetek nagy többségében a hangüzenet-beszédatbázisban megtalálható a szintetizálni kívánt szó, szófűzer, esetleg teljes mondat a megfelelő prozódiai szerkezettel. Ebből következik, hogy az összefűzési pontok szinte mindegyike szóhatárra esik a szintézis során.

Magyar és angol nyelvű bemondásokra is fel kellett készíteni a rendszert. Ezért elkészítettük a magyar rendszernek megfelelő angol hangfelvételeket és azok feldolgozását is.

A válogató és összefűző program

A szöveg-beszéd átalakítást vezérlő szoftver a BME TMIT-en készült, működését az alábbi folyamatábra mutatja (3. ábra).



3. ábra

A pályaudvari magyar-angol nyelvű hangos utas tájékoztató működési rendszere

A bemondó kiválasztása

A beszélő kiválasztása összetett folyamat, több ellentétes szempontnak kellett megfelelnie a kiválasztott hangnak. A szempontokat három fő csoportra oszthatjuk: beszédtechnológiai, szubjektív érzeti és szervezési vonatkozások.

Beszédtechnológia szempontból lényeges, hogy a hangüzenet-beszédatbázis készítésénél használt gépi elemző eljárások jól tőrjék az alkalmazott hangot, a méréseket a lehető legpontosabban el lehessen végezni. Például jól

artikulált, tiszta ejtés esetén a beszédfelismerő kényszerített módban történő futtatása pontosabb hanghatárjelölést ad, amely csökkenti az utólagos kézi feldolgozás ráfordításigényét. Ugyanakkor az irreguláris zöngéképzés gyakori előfordulása akadályozza az alapfrekvencia-menetek pontos meghatározását, tehát kerülni kell az ilyen beszédet. Az egyenletes kimeneti beszédminőség elérése miatt szükséges az egyenletes artikulációs sebesség akár több óráos felvétel esetén is. A nagy pályaudvarok visszhangos akusztikai terei miatt a női hang preferált, a gyorsabb visszhanglecsengés miatt.

Mivel a hangosbemondó beszédét sok ember hallgatja, ezért szubjektív érzeti szempontból is elfogadottnak kell lennie a beszélőnek. A beszédnek kellemesnek, bizalmat árasztónak és jól érthetőnek kell lennie.

Mivel a beszédhangra hosszabb távon (több évig) is szükség lehet, ezért a beszélővel szemben elvárás, hogy olyan életkori szakaszban legyen, amikor várhatóan a beszédhangja csak kis mértékben változik az idő során. A korpuzsalapú beszéd szintetizátor nagyméretű adatbázisa miatt szükséges, hogy a beszélő több órán vagy napon keresztül is egyenletes minőségben tudjon beszélni, rá tudjon hangolódni egy korábban ejtett hangzásra.

A fenti szempontokat összegezve olyan beszélők jöhetnek számításba, akik nagy tapasztalattal rendelkező képzett beszélők, ilyenek például a színészek és a rádióbemondók. Esetünkben további fontos követelmény volt a többynnyelvűség az állomásnevek kiejtése szempontjából is és az angol bemondások miatt is. A választás a Kossuth rádió női bemondójára, Mátyus Katira esett.

A fejlesztés lépései

Az ismertetett korpuzsalapú TTS fejlesztése 4 lépésből állt.

1. lépés. A felolvasandó szöveg megtervezése; a felolvasás megtervezése és végrehajtása; az elkészített hangfelvételek annotálása, címkézése és egyéb feldolgozása; a keresőszoftver fejlesztése, a rendszer elkészítése.

2. lépés. A működés laboratóriumi tesztelése, vagyis a keresőalgoritmus vizsgálata (a bemenő szöveg alapján milyen elemeket jelöl ki összeillesztésre, és mely szövegkörnyezetekből, hány mondatból stb.). Több száz mondat meghallgatása és a talált hangzási hibák értékelése, feldolgozása, kijavítása.

3. lépés. Rendszerbe illesztés és éles tesztelés. A valós környezetben történő tesztelés során ellenőriztük és beállítottuk a pályaudvari hangrendszer optimális átviteli karakterisztikáját, valamint a vezérlőegység működését (helyes adatokat ad-e). Az „éles” szövegek helyes felolvasását meghallgatással ellenőriztük a pályaudvaron. A tesztüzem során sablonokat alakítottunk ki a gyakorlati szövegezési munka segítésére, valamint teszteltük a felügyeleti rendszert is (az ellenőrző személy munkáját, a meghallgathatóság működését a rendszert működtető adatfelvivő segítésére stb.). Nyomkövető fájlok generálását is biztosítottuk (log), ezekkel lehet az utólagos offline ellenőrzéseket elvégezni.

4. lépés. Éles üzem. Egy ilyen komplex rendszer üzemeltetéséhez, napi szintű jó működéséhez folyamatos nyelvi támogatás kell mind az üzemeltető, mind a fejlesztő részéről (idegen állomásnevek kiejtésének meghatározása, leírása hangszinten, ugyanazon kifejezések használata adott tartalom megfogalmazásakor stb.). A vasúti régiók állomási szövegfogalmazási formáival szemben korábban nem volt semmi követelmény, mivel a leadott szöveget egyedileg készítették el, azaz egy bemondó felolvasta a mondatot. Így mindegy volt, hogy milyen szavakat, szókapcsolatokat használt a szöveg megfogalmazója. Az új rendszerben a nyelvi támogatással segítjük az egységes nyelvezet kialakítását, hiszen ez garantálja a kiváló hangminőséget.

Eredmények, a rendszer tulajdonságai

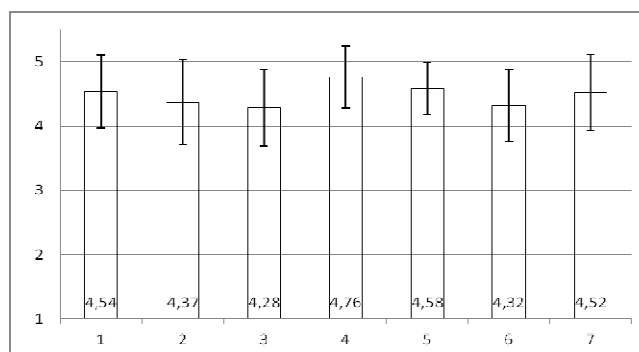
A megvalósított rendszer kielégíti a követelményekben megadott szempontokat: az előre kiválasztott női hangon szól; jó a hangminősége (felismerhető az ismert rádióbemondó hangja); sokkal rugalmasabban lehet hangüzeneteket előállítani, mint a korábbi megoldással; dinamikus rendszerként üzemeltethető; hirtelen igény esetén is lehet azonnal új hangüzeneteket generálni (a témakörön belül); magyar és angol nyelven ugyanazon hang szól; magyar-angol üzenetrészek kombináltan is megszólaltathatók (az angol közleményben a magyar állomásnevet nem angolul, hanem magyar kiejtéssel mondja, a román román kiejtéssel, a németet némettel, stb.); távolról is vezérelhető; teljes körű, országosan is egységes megoldássá lehet bővíteni (ugyanazon hanggal); hosszú időre megoldja a vasútállomásokon a hangos üzenetek előállítását (nem kell hangfelvételt csinálni, ha a jövőben a HMM- technológia (Tóth–Németh 2010) támogatását is hozzácsatoljuk); más vasúti felhasználásokba is rugalmasan beépíthető (fedélzeti hang, tájékoztató infó, oszlopok hangja, szolgálati belső rendszerek stb.). Mindezen előnyök mellett az üzemeltetőnek is illeszkednie kell a rendszerhez, újszerű kezelést kell megtanulnia, a korábban berögződött üzemeltetés nem folytatható: precíz előkészítő munkát kíván, mert online működik.

Kiértékelés

Az elkészített rendszert internetes meghallgatásos teszttel is értékeltettük. A tesztalanyoknak 7, az elkészült beszédsszintetizátor témakörébe eső mondatot kellett értékelniük a kiírási követelmények 5-ös skálája szerint (lásd korábban). A legjobb osztályzatot az 5 jelenti. A teszt elején a kísérleti alanyok egy ismertetőt olvashattak a kísérletről és annak menetéről, majd néhány adat (nem, kor, eszköz, beszédtechnológiai ismeretek) megadását kértük tőlük.

A tesztben 50 fő vett részt, ebből 29 férfi és 21 nő; 21 és 74 év közöttiek, átlagéletkoruk pedig 40,58 év volt. Az 50 főből 7 volt olyan, aki a teszt elején feltett kérdések alapján beszédtechnológiai szakértőnek vallotta magát. Az elkészült MOS (Mean Opinion Score) teszt eredményeinek kiértékelése szerint (4. ábra) mindegyik mondatra átlagosan legalább 4,28-as értékelést adtak

az adatközlők. A legjobb értékelés átlaga 4,76. A 7 mondatra az együttes értékelés átlaga 4,48. Ez az eredmény áttekintően azt jelenti, hogy a gépi beszéd hangminősége ebben az utastájékoztató rendszerben nagyon jó.



4. ábra

A vasúti hangos utastájékoztató rendszer hangjának MOS osztályozati átlagai az egyes mondatokra

Összefoglalás

A pályaudvari hangos utastájékoztatót kiszolgáló legújabb magyar fejlesztést mutattuk be. Az új rendszer rugalmasabb kezelést biztosít, mint a korábbi megoldások, olcsóbb az üzemeltetése, ugyanakkor stabilabban szolgáltat jól érthető, szívesen hallgatott hangminőséget. A rendszer hangja meghallgatható a Budapesti Keleti pályaudvaron és számos más MÁV állomáson.

Irodalom

- Fék, Márk – Pesti, Péter – Németh, Géza – Zainkó, Csaba – Olasz, Gábor 2006. Corpus-based unit selection TTS for Hungarian. In *Text, Speech and Dialogue*. Springer Berlin, Heidelberg. 367–373.
- Klabbers, Esther 1997. High-quality speech output generation through advanced phrase concatenation. In: *Proceedings of the COST Workshop on Speech Technology in the Public Telephone Network: Where are we today?* Rhodes, Greece. 85–88.
- Tóth, Bálint – Németh, Géza 2010. Improvements of Hungarian Hidden Markov Model-based text-to-speech synthesis. *Acta Cybernetica* 19/4. 715–731.
- Zhenli, Yu – Hong, Wu – Mengchu, Wu – Guilin, Chen 2012. Speech corpus script design for TTS system applied on railway passenger service information broadcasting. In: *Proceedings of Oriental COCOSA*. 97–100.
- Zainkó Csaba – Németh Géza 2010. Vasútállomási utastájékoztató. In Németh Géza – Olasz Gábor (szerk.): *A magyar beszéd*. Akadémiai Kiadó, Budapest. 579.

Zainkó Csaba 2010. Elemkiválasztás-alapú szövegfeldolvasó. In Németh Géza – Olasz György (szerk.): *A magyar beszéd. Beszédkutatás, beszédtechnológia, beszédinformatika rendszerek*. Akadémiai Kiadó, Budapest. 505–512.

A kutatást az EITKIC_12-1-2012-0001 projekt támogatta, amely a Magyar Kormány támogatásával, a Nemzeti Fejlesztési Ügynökség kezelésében, a Kutatási és Technológiai Innovációs Alap finanszírozásával és az EIT ICT Labs Magyar Nemzeti Társult Csomópont közreműködésével valósul meg. (www.ictlabs.elte.hu)

ANNOTATION PRO: ÚJ ESZKÖZ A BESZÉD NYELVI ÉS PARANYELVI SAJÁTOSSÁGAINAK CÍMKÉZÉSÉRE ÉS ELEMZÉSÉRE

Katarzyna Klessa – Markó Alexandra

Bevezetés

Annotáláson azt a folyamatot értjük, amely során a beszéd meghatározott részeihez jelölőelemeket (ang. *label*, *tag*) rendelünk. Az elemzés céljától függően ezek a jelölőelemek bármilyen beszédszeletre vonatkozhatnak a fonetikai szegmentumoktól kezdve a komplex frázisokig vagy akár még nagyobb egységekig. A jelölőelemek általában egy explicit, zárt elemtár tagjai, mint például a beszédhangok kategóriáinak jelölői (magánhangzók, mássalhangzók, fonetikus ábécé, intonációs jelölők stb.) vagy a nyelvészeti leírások jelölői (mondatrészjelölők, morfológiai kommentárok). Ennek ellenére vannak olyan kutatási célok, amelyek nagyobb rugalmasságot kívánnak meg, és ahol a tulajdonságok mértékét (egy kontinuumon elfoglalt helyét) jelöljük, vagy olyan kategóriákat határozzunk meg, amelyeknek elmosódottak a határaik. Ilyen például a beszéd emocionális aspektusainak a jelölése, ahol bizonyos köztes affektív állapotokat kívánunk megkülönböztetni a szélső értékek, például az öröm és a szomorúság között. Az ilyen köztes állapotok számát az annotálók önkényesen határozzák meg. Ugyancsak problémaként merülhet fel, hogy néhány jelölő két- vagy többdimenziós, vagyis az értéke is csak többdimenziós térben adható meg. Ezért akár két vagy több értéket (jelölőt, taget) is tartalmazhatnak, amelyek különböző címkesorokon/rétegeken (ang. *layer*; a Praatban *tier*, vö. Boersma–Weenink 2014) jelennek meg. Ezért szükséges a jelölő(elem) (ang. *tag*) és a címke (ang. *label*) fogalmi elkülönítése, amit Popescu-Belis (2005) javasolt.

Több, a beszéd annotálására szolgáló program (vö. pl. Garg et al. 2004) sok lehetőséget kínál, de ez a bőség olykor kevésbé áttekinthető felhasználói felületet és komplex működést eredményez. Néhányat ezek közül eredetileg eszközfonetikai elemzésekre hoztak létre, és csak kevés támogatja közvetlenül a nem kategorikus vagy a komplex annotációt. Az Annotation Pro kifejelesztésének fő motivációját egy olyan szoftver iránti igény adta, amely többszintű annotációt tesz lehetővé, különféle kategorikus és nem kategorikus minősítésekkel. A szoftver mind beszédakusztikai, mind percepció alapú elemzésekre lehetőséget ad. Mivel minden típusú bemenetet egy annotációs fájlban tárol (.ant), könnyen kombinálhatók a kétféle elemzés adatai, így hozzáférhetővé válnak a köztük lévő lehetséges kapcsolatok és összefüggések.

Kezdeti megfontolások és dizájn

A paranyelvi és más nem kategorikus jegyek kihívást jelentenek a beszédadatok annotálásában – mind a szoftver, mind a humán címkéző szempontjából. Az annotációs tér meghatározásának módja nagyban befolyásolja a kutatási eredményeket. Annak is jelentősége van, hogy milyen típusú skálát választunk egy adott dimenzióra (lineáris, logaritmikus stb.). A paranyelvi jegyeket gyakran nehéz egyértelműen, világos és hozzáférhető terminusokkal definiálni. Ha verbális jelölőket (tag) alkalmazunk (pl. emocionális kategóriák neveit, mint *ingerült*, *aggódó*), ezeknek a címkéző általi értelmezését erősen befolyásolhatja e szavak mindennapi használata. Az ilyen és ehhez hasonló problémák csak részben oldhatók meg, és a megoldások általában egy adott adattípusra vagy egy specifikus tudományos kutatásra alkalmazhatók.

Az itt bemutatott programot a beszédcímkézés széleskörű alkalmazásaira fejlesztették ki, beleértve a szigorúan technológiaiakat (pl. beszéd-szintézis, automatikus beszélő- vagy beszédfelismerés) éppúgy, mint az alapvető fonetikai kutatásokat vagy a kommunikáció pszichológiáját. Ennek megfelelően elengedhetetlennek bizonyultak az alábbi funkciók és lehetőségek:

- egyszerű és felhasználóbarát kezelőfelület, könnyű telepítés és konfigurálás;
- többszintű, szinkronizált annotáció precíz határellhelyezéssel;
- a jegytérrel különféle, állítható grafikai reprezentációi, valamint feltölthető képekkel elérhető skálák;
- lehetőség arra, hogy a felhasználó a saját – képként reprezentált – tereit és skáláit használja;
- komplex címkék (tagek) használata (pl. kétdimenziós jegyek);
- plugin architektúra a program funkcióinak kiterjesztésére a felhasználó igényei szerint;
- átjárhatóság: adatok importálása és exportálása más programokból/ba.

Egy, a paranyelvi jegyek annotációjára alkalmas szoftver percepcióalapú kutatások kísérleti eszközeként is szolgálhat. A felülről lefelé haladó megközelítés, amely előre meghatározott kategóriákkal dolgozik az annotálásban, a

címkézőkét és a kutatókat is vakká teszi az új, korábban fel nem tárt jelenségekre. Ezzel szemben az itt bemutatott eszköz több rugalmasságot enged meg a címkézőknek, és azáltal, hogy folytonos tér áll a rendelkezésükre, új felfedezések tehetők, vagy új kategorizáció születhet meg.

Implementáció és architektúra. Pluginok

Az Annotation Pro C# programnyelv használatával, Visual Studio programozási környezetben készült, és (a jelen formájában) Windows operációs rendszeren futtatható. Más platformokhoz (pl. Linux) illesztése tervben van. A fő cél az volt, hogy olyan általánosan használható annotációszoftver jöjjön létre, amely különféle projektumokban alkalmazható, mind beszélt, mind írott (pl. morfológiai kommentár) forrásokon. Másrészt az architektúra legyen kiterjeszthető és rugalmas, lehetővé téve a felhasználóspecifikus igényeknek való megfelelést. Ez a plugintechnológia alkalmazásával történik meg, amely a felhasználó számára lehetővé teszi, hogy saját funkciókat adjon hozzá a program főmenüjéhez.

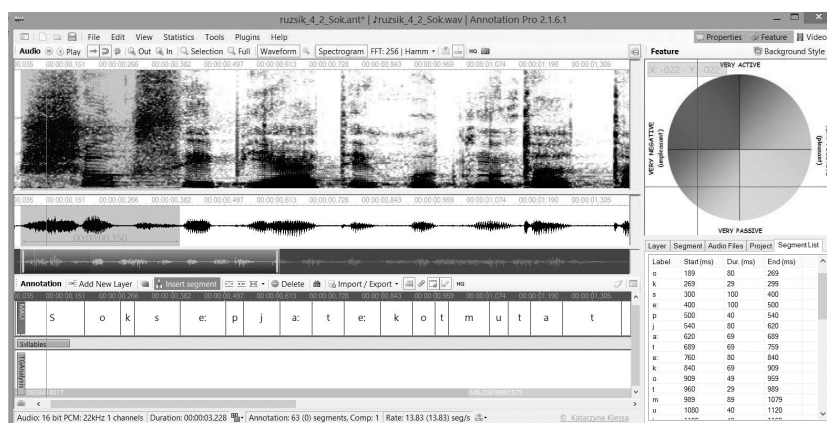
A plugintechnológia alkalmazása természetesen következik a fő cél megvalósításából: csak az általános, a legtöbb felhasználó által használt lehetőségek vannak fixen beépítve, míg az összes olyan funkció, amely inkább projektum- vagy felhasználóspecifikus, a pluginmenün keresztül érhető el. A C# nyelvben jártas felhasználó létrehozhat bármilyen új funkciót a szoftver működésének kiterjesztésére (pl. beszédelemzés, automatikus jegykinyerés a beszédjelből, szinkronizálás). Inicializáláskor a program beolvassa a felhasználó Documents/Annotation Pro/Plugins könyvtárának tartalmát, és frissíti a pluginok listáját ennek a könyvtárnak a tartalma alapján. A pluginok standard C# formátumot (*.cs) és megfelelő szerkezetet kívánnak meg, erre egy példa plugin fájl található az Annotation Pro Plugins könyvtárban. A plugin fájl (*.cs) fordítása és végrehajtása is futásidőben történik. A felhasználó korlátlanul hozzáférhet a felülethez, címkézhet rétegeket és adatokat (az architektúra implementációjának további részleteiről: Klessa et al. 2013; példaalkalmazások a beépített lehetőségekhez, valamint elemzési célú pluginarchitektúra az időzítés jelenségéhez társalgási, illetve felolvasott beszédben: Klessa–Gibbon 2014; Yu et al. 2014; Karpiński 2014).

Az annotálásra és a percepció tesztelésre szolgáló felület

A „hagyományos” többretegű annotációs felületektől eltérően (ahol mind a hangszínek, mind a hullámforma megjelenik, mint az 1. ábrán is látható), itt működik egy univerzális grafikus vezérlő, amely lehetővé teszi, hogy a jegytér különböző grafikus reprezentációit használjuk az annotáció és a percepció tesztek alapjaként (lásd az ábra jobb felső sarkában). A grafikus reprezentációk a programba beépített képek listájából is kiválaszthatók, vagy a felhasználó is megtervezheti őket. Ez azt jelenti, hogy bármely kép használható grafikus elemként. A grafikus vezérlő képéhez tartozó síkot a szoftver koor-

dináta-rendszerként értelmezi. Amikor a címkéző a képre kattint, az adott pont koordinátái eltárolódnak, és megjelennek pontként a képen, valamint számokként a hozzá tartozó címkesorban. Ennek eredményeként a koordináták halmazain különféle elemzések futtathatók, pl. klaszteranalízis (más grafikus reprezentációkra lásd pl. Klessa et al 2012, ahol a FEELTRACE értelemezővel (Cowie et al. 2000) történt összevetés is olvasható).

Az annotációs felület percepciós tesztekhez is használható: tesztfuttatási módban lehet kísérletet tervezni. Lehetőség van a résztvevők adatainak kezelésére (a résztvevő neve vagy azonosítója, életkor, nem, nyelvjárás stb.). A percepciós teszt beállításai is rugalmasak, az adott feladatra szabhatók. A kísérletvezető meghatározhatja a lehetséges válaszok számát, a stimulusok sorrendjét, az újra meghallgatás lehetőségét stb. Az eredeti fájlnev megjeleníthető vagy el is rejthető a teszt ideje alatt. A percepciós tesztek eredményei .csv fájlokban tárolódnak, velük együtt minden információ a kísérleti személy teszt közbeni cselekedeteiről (pl. a meghallgatások száma).



1. ábra
Az Annotation Pro felhasználói felülete

Átjárhatóság és fájlformátumok

Az Annotation Pro annotációs fájllai (.ant) XML formátumúak, és a program más eszközökkel előállított XML-fájlokat is kezel (Klessa et al. 2013). Ezen kívül az Annotation Pro importál fájlokat a következő formátumokból: a Praat TextGrid-fájlja (Boersma–Weenink 2014), a Transcriber .trs fájlja (Barras et al. 2001), BLF (Breuer–Hess 2010), .txt (a forrás szövegfájl minden sorát különálló szegmentumként importálja) és .csv (konfigurálható import, beleértve a Wavesurfer LAB-ját (Sjölander–Beskow 2000)).

Konklúzió és tervek

Az Annotation Prót eddig különféle kutatási projektumokban alkalmazták beszédadatok lejegyzésére és címkézésére. Ezek között szerepelt emocionális beszéd percepció és produkció elemzése, valamint időzítési sajátosságok elemzése mind társalgási, mind felolvasott beszédben. Ugyancsak használták szűkebb körben beszélt nyelvek korpuszainak annotálására és lejegyzésére nem standard betűtípusokkal és morfológiai kommentárokkal. Az eddigi visszajelzések alapján a szoftver elve világos, a felhasználói felület könnyen kezelhető és rugalmas.

Az Annotation Pro további tesztelése és új pluginokkal való kiterjesztése folyamatban van. Mind az import/export és az elemzés lehetőségeit továbbfejlesztik a nagyobb átjárhatóság érdekében (pl. SPPAS, Bigi–Hirst 2013). Egyebek mellett tervben van a videóannotáció és a több-beszélős felvételek kezelése is.

Az Annotation Pro kutatási célokra szabadon elérhető: annotationpro.org (kapcsolat: klessa@amu.edu.pl).

Irodalom

- Barras, Claude – Geoffrois, Edouard – Wu, Zhibiao – Liberman, Mark 2001. Transcriber: Development and use of a tool for assisting speech corpora production. *Speech Communication* 33/1–2. 5–22.
- Bigi, Brigitte. 2012. SPPAS: a tool for the phonetic segmentation of speech. In: *Language Resource and Evaluation Conference*. Istanbul, Turkey. 1748–1755.
- Boersma, Paul – Weenink, David 2014. Praat: doing phonetics by computer (Version 5.4) [Computer program]. <http://www.praat.org/>
- Breuer, Stefan – Hess, Wolfgang 2010. The Bonn Open Synthesis System 3. *International Journal of Speech Technology* 13/2. 75–84.
- Cowie, Roddy – Douglas-Cowie, Ellen – Savvidou, Susie – McMahon, Edelle – Sawey, Martin – Schröder, Marc 2000. FEELTRACE: An instrument for recording perceived emotion in real time. In: *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*. 19–24.
- Garg, Saurabh – Martinovski, Bilyana – Robinson, Susan – Stephan, Jens – Tetreault, Joel – Traum, David R. 2004. Evaluation of transcription and annotation tools for a multi-modal, multi-party dialogue corpus. In: *Proceedings of the LREC*. 2163–2166.
- Gibbon, Dafydd 2013. TGA: a web tool for Time Group Analysis. In Hirst, Daniel – Bigi, Brigitte (eds.): *Proceedings of the Tools and Resources for the Analysis of Speech Prosody (TRASP) Workshop*. Aix en Provence. 66–69.
- Karpiński, Maciej – Klessa, Katarzyna – Czoska, Agnieszka 2014. Local and global convergence in the temporal domain in Polish task-oriented dialogue. In Campbell, Nick – Gibbon, Dafydd – Hirst, Daniel (eds.): *Proceedings of the 7th Speech Prosody Conference*. Dublin, Ireland. 743–747.
- Klessa, Katarzyna – Gibbon, Dafydd 2014. Annotation Pro + TGA: automation of speech timing analysis. In: *Proceedings of the 9th Language Resources and Evaluation Conference*. Reykjavik, Iceland. 1499–1505.

- Klessa, Katarzyna – Karpiński, Maciej – Wagner, Agnieszka 2013. Annotation Pro – a new software tool for annotation of linguistic and paralinguistic features. In Hirst, Daniel – Bigi, Brigitte (eds.): *Proceedings of the Tools and Resources for the Analysis of Speech Prosody (TRASP) Workshop*. Aix en Provence. 51–54.
- Klessa, Katarzyna – Wagner, Agnieszka – Oleśkiewicz-Popiel, Magdalena 2012. Using Paralingua database for investigation of affective states and paralinguistic features. *Speech and Language Technology* 14–15. 71–91.
- Popescu-Belis, Andrei 2005. Dialogue Acts: One or More Dimensions? *ISSCO Working Paper 62*. University of Geneva, Geneva.
- Sjölander, Kåre – Beskow, Jonas 2000. WaveSurfer – an open source speech tool. In: *Proceedings of 6th ICSLP Conference 2000*. Vol. 4. Beijing. 464–467.
- Yu, Jue – Gibbon, Dafydd – Klessa, Katarzyna 2014. Computational annotation-mining of syllable durations in speech varieties. In Campbell, Nick – Gibbon, Dafydd – Hirst, Daniel (eds.): *Proceedings of the 7th Speech Prosody Conference*. Dublin, Ireland. 443–447.

„NYELVBOTLÁS”-KORPUSZ

11. rész

Az első magyar valós idejű, jegyzeteléses technikával gyűjtött megakadás-jelenség-korpuszt 2004 óta adjuk közre folyóiratunkban:

Beszéd kutatás 2004. 19–186.: 5139 adat;

Beszéd kutatás 2005. 145–173.: 761 adat;

Beszéd kutatás 2006. 231–247.: 388 adat;

Beszéd kutatás 2007. 187–198.: 244 adat;

Beszéd kutatás 2008. 221–239.: 444 adat;

Beszéd kutatás 2009. 257–267.: 176 adat;

Beszéd kutatás 2010. 283–291.: 179 adat;

Beszéd kutatás 2011. 149–165.: 429 adat;

Beszéd kutatás 2012. 301–313.: 298 adat;

Beszéd kutatás 2014. 253–276.: 604 adat.

Eddig tehát összesen 8662 megakadásjelenséget elemeztünk, ebben a részben pedig újabb 193 tételt adunk közre.

Ertelemszerűen ez a korpusz a hiba típusúak gyűjteménye. A kezdetektől számos, különféle szempontú tanulmány, szakdolgozat, disszertáció készült ebben a témakörben. Mindez a tapasztalat oda vezetett, hogy a „Nyelvbotlás”-korpusz osztályozási kategóriáit 2009-től egyszerűsítsük. Az alapvető ok az volt, hogy a korpuszt használóknak nagyobb szabadságot biztosítsunk egy adott jelenség mélyreható elemzéséhez.

Az egyszerűsített osztályozás elve az, hogy minden esetben a felszínen tapasztalható jelenséget soroljuk kategóriákba, a finomelemzések az adott jelenség kutatójára várnak. A fő kategóriák és meghatározásuk a következők:

1. Téves szó

A megfelelő szó helyett egy másik szó megjelenése a felszínen. Ez a kategória tartalmazza a freudi elszólásokat és a malapropizmusokat is, amelyek további elemzésekkel különíthetők el.

2. Grammatikai hiba

A köznyelvi normának ellentmondó morfológiai/szintaktikai struktúra.

3. Kontamináció

Két nyelvi jel (szó, szerkezet) összeolvadása, vegyülése.

4. Téves kezdés

Beszédhang vagy beszédhangsorozat ejtése, amely azonban nem szó.

5. Nyelvem hegyén van jelenség

A szándékolt szó jelentésének és morfológiai struktúrájának ismerete, de a fonológiai/fonetikai szerkezet kiejtésének (részleges) gátoltsága.

6. Perszeveráció

Egy kiejtett nyelvi elem szándéktalan, ismételt megjelenése a közlés egy későbbi időpontjában.

7. Anticipáció

Egy szándékozott nyelvi elem megjelenése a közlés egy korábbi időpontjában is.

8. Metatézis

Nyelvi elemek sorrendjének felcserélődése.

9. Egyszerű nyelvbottlás

A beszédtervezés artikulációs tervezési szintjén, ill. az artikulációban létrejött hiba, amely semmilyen egyéb okkal nem magyarázható. Altípusai: a) betoldás, b) helyettesítés, c) kiesés

10. Több típusba sorolható jelenségek

Azok az adatok kerülnek ebbe a kategóriákba, amelyek létrejöttében a felszíni elemzés alapján több rejtetten működő folyamat hibája is feltételezhető.

1. Téves szó

Megakadás	Szándékolt közlés
A hollywoodi filmstúdiókba is betekintést tett.	betekintést nyert
A kísértésben □ kísérlésben vizsgálták azt is, hogy...	
A kultúra hagyománya a sok alkoholfogyasztása.	a sok alkoholfogyasztó
A történetben a forduló □ fordulatot az hozza, hogy felfedeznek egy időgépet.	
A versenyzők már 17 órától elkezdi a szónok írását.	szöveg
A virslit vegyük már ki abból a tányérból, vagyis nejlomból.	
Akik ide, az Eötvösbe szeretnének lenni...	jönni
...az időjárás-hirdetés vagyis jelentés...	
Azt hittem, a kékhajú lesz az.	kékruhás
Elkezdték keresgélni a szövegek iránt.	után
Érettségiztettem a diákom.	érvényesítettem
Fájdalmas jövő áll számára.	előtte
Gyerekkorom óta foglalkozom énekességgel □ énekléssel.	
Ha nem pont így mondtam, akkor elnézést kívánok.	elnézést kérek
Hiába el tud jönni ő is.	hátha
Itt nem jár vonat vagy na villamos.	
Jön a vágány.	villamos

Megakadás	Szándékolt közlés
Könnyű az esély arra, hogy...	nagy az esély
Lemegyek a bárba vagyis Sparba.	
Mennyi a temetkezési idő?	tűrelmi
Nem az ő monológja volt.	monogramja
Nincsen hangszóró, így sajnos nem tudok ve-	
títeni.	lejátszani
Oda kell rá vigyázni.	oda kell rá figyelni
...tehát ez viszonylag kisméretű □ mértékű	
anyag.	
Telefonba mondta be a szöveget.	mikrofonba
Valamennyi időnek fel kell kelnie...	el kell tennie
Válassza □ fel a feladat megoldását.	vázolja
Vedd fel a papucsod vagyis csizmád!	

2. Grammatikai hiba

Megakadás	Szándékolt közlés
A képregénnyel később kerültem kapcsolat-	
ban.	kapcsolatba
A kínai felküld □ a kínaiak felküldik az el-	
ső műholdjukat a Holdra.	
A táblába írtam a vázlatot, és nem másolták	táblára
le.	
Az állami gondozásban élő gyerekeket érint	érinti
a legtöbbször.	
...az életkorok □ az életkor	
előrehaladtával...	
Az mennyire vegyes csapatok?	azok
Az összesek eljöttek □ az összes eljött.	
Azért ezeket a gyerekekhez kell pszicholó-	ezekhez
gus.	
...egészséges magzat az anya hasuk □ hasá-	
ban.	
Egy fiút □ egy fiú van, aki nincs felmentve	
tesiből.	
Egyrészt a haja □ hajuk balzsamozott az	
emósoknak...	
Először te mondom el, aztán én.	mondod
Esélye van elindult □ elindulni.	
Ez a felmentések számából lehet mérni.	ezt
Ez már az utolsó tanulmányból származnak.	ezek
Ezért tesznek bele töltelékszavak □ szót.	

Megakadás	Szándékolt közlés
Felébredtem nyolctól <input type="checkbox"/> nyolckor.	
Ha meghallgatjuk <input type="checkbox"/> meghallgatunk egy mondatot...	
Itt lakunk, közel a ligetbe.	ligethez
Itt nem vannak helyek...	nincsenek
...jelentkeztek <input type="checkbox"/> jelentkezhettek a hte@hte.ro webcímen.	
Kialakulnak a nagy zenekar és a nagy bandák.	zenekarok
Kilencvenben annyi, mint legutóbb volt.	kilencvenszer
Megcsinálták a feladat.	feladatot
Nem ülnek le tanulni, ezért a szókinsük nehezebben fejlődik.	szókinsük
Nincsenek szükségeim indokokra.	nincs szükségem
Piacra dobunk egy toastkenyeret, ami szintén rajta lesz a címke.	amin
Sikerült eszembe jutnom...	jutnia
Sokan a szövegértésben vannak nehézségeik.	sokaknak
Szenvedős arccal vágok mindig.	arcot
Te is jól érezted magam?	magad
Úgyis van hibák az én beszédemben.	hiba

3. Kontamináció

Megakadás	Szándékolt közlés
Beüvöltöttem a kezem.	beütöttem × üvöltöttem
...borosz <input type="checkbox"/> boldog oroszt még nem láttam arra.	
Ebből fadódik az is, hogy...	fakad × adódik
Ideadnád a fejesemet?	fejhallgató × füles
Nem hiszem el, hogy ennyire nem vagy komolytalan.	ennyire komolytalan vagy × nem vagy komoly
Töltött fel fészkepeket.	fényképeket × facebook
...új fájta <input type="checkbox"/> új fajta csomagolása van.	fajta × fële
Velletted lehetek.	veled × melletted

4. Téves kezdés

Megakadás
A görög kлом <input type="checkbox"/> komplexitás jelentése...
A harmadik lemezüket ne <input type="checkbox"/> november negyedikén adják ki.
A Havanna úti leke <input type="checkbox"/> lakótelepen működik.

Megakadás

A szín ☐ a szving népszerű muzsika.
 A tornateremben k ☐ többek között fellép az iskolakórus is.
 Átlagosan tizenké ☐ tizenhét alkalom rémlik...
 És hogy mer ☐ miért mentették föl testnevelésből?
 Ez kláp ☐ kleptománia, ami nem kémiai függőség.
 Fontos, hogy tudjanak elt ☐ elhelyezkedni.
 Három vagy négy tom ☐ trombita játszik a zenekarban.
 Hé ☐ hé ☐ hé ☐ na, tizenhétszer annyi lett.
 ...illetve alkoholos bes ☐ alkoholos befolyásoltság alatt.
 Ilyen például a félintve ☐ félinterpretatív beszéd...
 Ilyesmivel foglalkozo ☐ szeretnék foglalkozni.
 Jó, majd a besz ☐ a végén besz ☐ a végén beszélek arról.
 Kicsi a szócsincs ☐ szókincse.
 Kimulattátok magalo ☐ magatokat?
 Korom ☐ Kolumbusz Kristóf...
 Különge ☐ különleges meglepetéssel fog várni.
 Különösen grí ☐ krízis közepén nehéz együttműködni.
 Legközelebbi nyelvrokok ☐ nyelvrokonokról lehet kérdezni.
 Luca napi népszopá ☐ népszokás...
 Luxushaj ☐ luxusautó miatt is vitatkoznak.
 Majd próbálok egyenesen ravi ☐ lavírozni.
 Már országszerte rez ☐ rendezvényeket szerveznek.
 Már várjuk a lej ☐ jelentkezőket.
 Megnézzük a ne ☐ nagy nemzetközi szavazatállásokat.
 Milyen sz ☐ célból jöttetek el a nyílt gyűlésre?
 Mindenki képes lenne tökén ☐ tökéletesen leutánozni azt a hangsort.
 Mit ünneplünk püs ☐ pünkösdkor?
 Po ☐ percre pontosan mit kell tanítani?
 Soha nem érték el egy színv ☐ szimfonikus zenekar méreteit.
 Sok szike ☐ sikert nektek!
 Sose te ☐ talállok bele.
 Tok ☐ tökéletes előadás volt
 ...valamint a valá ☐ valódi spontán beszéd.
 Viszont az lett a kövekezm ☐ vagyis az a következtetés, hogy nem oldható meg.

5. „Nyelvem hegyén van” jelenség**Megakadás****Szándékolt közlés**

A hangszer ☐ hangszerkezet ☐ hangszer-
 szekciói kialakulnak a zenekaroknak.
 Konfide ☐ konfirma ☐ szerettem volna. konfirmálni

Megakadás	Szándékolt közlés
Szükség lesz a vonat <input type="checkbox"/> bevonat <input type="checkbox"/> kivonat <input type="checkbox"/> na az anyakönyvi vonat <input type="checkbox"/> a születési anyakönyvi kivonatodra.	

6. Perszeveráció

Megakadás	Szándékolt közlés
A zenekar egyik tagja <input type="checkbox"/> tagja kilépett. Enyém, tiém... Indul az ifjúsági misor <input type="checkbox"/> műsor. Leszűkülnek az improvizációs lehetséges <input type="checkbox"/> lehetőségek. Mekkora diszkóól van itt! Mondom, én briliáns biliárdos vagyok. Sanders sztílus <input type="checkbox"/> stílustipológiája... Tegnap délután a körúron sétáltam. Tudja a megoldást a magyar igar <input type="checkbox"/> ipar problémáira. Utána ketten kaptak egy kerezdet. Várom a véré <input type="checkbox"/> a végét.	tiéd disznóól briliáns biliárdos körúton gerezdet

7. Anticipáció

Megakadás	Szándékolt közlés
Akkor kelten talizunk. A sztesztoszteron <input type="checkbox"/> tesztoszteron hormon... Az időgép visszaröpi őket az ezerkilenc- százhusas <input type="checkbox"/> húszas évekbe. Együnk palicsintát. Ez a trúra nagyon izgalmas lesz! Hatalmasat lépett erőre. Idelilek kikészültem. Itt egy méheckés ceruza. Jesz <input type="checkbox"/> lesz jelnyelvi tér a bölcsész napokon. Mem azért, mert érdekes <input type="checkbox"/> mert nem érde- kes... Most jól megfűszerelem a levest. Ott is egy jó poncertplakát. Ő is timikus mérnök. Puzzafutár <input type="checkbox"/> pizzafutár lett belőle. Tamács mit csinál?	kedden palacsintát túra hatalmasat lépett előre idegileg méheckés megfűszerezem koncertplakát tipikus Tamás

8. Metatézis**Megakadás**

A nyelven hegyém van.
 Add ide a szósórót!
 Barakcot szeretnék.
 Be kell adni lekötve.
 Csirkézett grillét kérek.
 Elmegyek szokszor ☐ sokszor.
 Finom a palács kirítva.
 Hogy csöppéntel ☐ csöppéntél bele a zenei életbe?
 Mostanában fedelékeny vagyok.
 Otthon kifejezem a beszínezést.

Szándékolt közlés

a nyelvem hegyén van
 sósórót
 barackot
 Le kell adni bekötve.
 grillezett csirkét
 kalács pirítva
 feledékeny
 befejezem a kiszínezést

9. Egyszerű nyelvbotlás**Megakadás**

A gépek mellett 45 hőmunkás ☐ hómunkás is dolgozik.
 A legtünőbb big bandekből emelnék ki párat.
 A madi ☐ mai adás a szving zenei műfaj köré szerveződött.
 A szavazatik alapján...
 A Szentéji ☐ Szentivánéji álom lesz a színházban.
 András és Lia ☐ Lilla öt éve járnak együtt.
 Barátnőm is próbálta, de hasztané...
 Demóanyag, amit az internetről lehetett ingyen ☐ ingyenesen letölteni.
 Egy évezredes sztereópiának próbáltak gátat vetni.
 Megbeszélték, hogy levelezőparterek ☐ partnerek lesznek.
 Mit cinálsz ☐ csinálsz vasárnap?
 Szonnyan halok.
 Vasárnap, azis ☐ vagyis február 24-én Akela-koncert lesz.
 ...hogyminden életkorban vegyük ☐ vegyük ☐ vegyünk ☐ na vegyünk föl, készítsünk felvételeket.

Szándékolt közlés

legfeltünőbb
 szavazataik alapján
 hasztalan
 sztereotípiának
 szomjan

10. Több típusba sorolható jelenségek

Megakadás	Szándékolt közlés
A férfi elvesztettnek <input type="checkbox"/> elvesztettnek érzi magát.	
A frizurád, remélem, kimarad.	megmarad
A mártiszu <input type="checkbox"/> márciusi szónokversenyt ötödik alkalommal rendezik meg.	
A szalma szósz nagyon finom lett.	salsa szósz
Az erdei táborba a részleget be kell fizetni.	részletet
Az összekötővel játszi <input type="checkbox"/> do <input type="checkbox"/> az összekötőnek dobja a labdát.	
Beesett vagy na, kiesett az ablakon.	
Ez a kísérlet jól sikerült.	kísérlet
Ez egy lapos <input type="checkbox"/> ööö <input type="checkbox"/> lápos terület.	
Ezen a szintén <input type="checkbox"/> szinten már semmi nem várható.	
Február elsejétől fog indulni egy táncfolyó <input type="checkbox"/> táncanfolyam.	
Gödöllőn most <input type="checkbox"/> most nem esett az eső.	Gödöllőn
Ha a pénztárca <input type="checkbox"/> ha a pénztárnál a pénztárcájukba jutnak <input type="checkbox"/> nyúlnak.	
Havi két hónap munka csak.	óra
Így lett a tizenkét <input type="checkbox"/> tizenkettedik...	
Ismertétek ezt a szervezetet a nyílt gyűlés <input type="checkbox"/> gyűl <input type="checkbox"/> gyűlés előtt?	
Itt megtanul <input type="checkbox"/> megtanítják őt bízni az emberekben.	
Kes <input type="checkbox"/> kes <input type="checkbox"/> kereskedő van a papíron	
Kiterítjük, hol lehet kapni.	kiderítjük
Kiveszett a lényeg az együttesükből <input type="checkbox"/> együttélésükből.	
Lángok mártalékává vált a ház.	martalékává
Ma a biztosítási köszvényekről lesz szó.	kötvények
Mennyit töltenek <input type="checkbox"/> költenek egy alkalommal?	
...mert mindig eljár sz buzizni <input type="checkbox"/> bulizni.	
Mi <input type="checkbox"/> ki volt a legsikeresebb tanítványod?	
Mi még többre vihettünk, előbbre juthattunk.	vihettük
Mi okoz gondot a tanulásban <input type="checkbox"/> vagy tanításban?	
Mindenem jól szerepelt.	mindenben

Megakadás	Szándékolt közlés
<p>Mindketten sokat fog érni egyszer. Mitől ilyen sörös <input type="checkbox"/> a párnád? Ne drummogj <input type="checkbox"/> durmolj <input type="checkbox"/> az órán! Néhány szó a várható ide <input type="checkbox"/> a várható hí <input type="checkbox"/> ő <input type="checkbox"/> időjárásról. ...nekem egyre újabban <input type="checkbox"/> újabban egyre inkább gyanús... Öntőberendezést <input type="checkbox"/> öntözőberendezést kell venni. Sosem hittem volna, hogy ennyit is kapha- tok róla <input type="checkbox"/> tőle.</p>	<p>mindkettő szörös brummogj</p>

A MAGYAR ÉS A NEMZETKÖZI TUDOMÁNYOS ÉLET ESEMÉNYEI

Gombocz Zoltán (1877–1935)



Idén 80 éve, hogy elhunyt Gombocz Zoltán, a magyar nyelvtudomány egyik jeles képviselője. Munkássága átívelt az egyes tudományterületeken: a fonetikát, a finnugor nyelvészetet, a grammatikát, a stilisztikát, a turkológiát és a nyelvtörténetet is számos új eredménnyel gazdagította. Gombocz korának kiemelkedő kutatója volt, aki kilépett a „puszta adatok” világából, és a hangsúlyt a „köztük lévő viszonyra helyezte” (Kiss 2011: 110).

Gombocz Zoltán Sopronban született, apja Gombocz Miklós evangélikus lelkész és középiskolai igazgató volt. Családjában tudós emberek vették körül, nagybátyja és unokatestvére nyelvész-filológus volt. Tanulmányait Sopronban, az evangélikus liceumban kezdte, majd Budapestre került, az Eötvös Kollégiumba, ahol magyar-francia szakra iratkozott be. A kollégium első nemzedékének tagjaként olyan emberek voltak társai, mint Horváth János, Szabó Dezső vagy Kodály Zoltán (Kiss 2011). Kollégiumi éve alatt szélesítette tudását, nagy hatással voltak rá nyelvészettel foglalkozó tanárai, Szinnyei József és Simonyi Zsigmond. Ennek jeles bizonyítéka, hogy Gombocz már harmadévesen publikált. 1900-ban franciatanári vizsgát tett, majd több tanulmányúton is járt: Lipcsében, Párizsban, Finnországban és Svédországban – elsősorban a fonetika tanulmányozása céljából (Kicsi 1997). 1906-ban magántanárrá minősítették, az általános fonetika és finnugor hangtan tárgykörében. Először a kolozsvári egyetemen nevezték ki tanárnak, majd 1921-ben visszatért Budapestre, és a budapesti egyetemen kapott állást. Mindezek mellett vendégtanárként a berlini egyetemen is tartott kurzusokat. 1927-ben nevezték ki az Eötvös Kollégium igazgatójává, ezt a tisztségét élete végéig betöltötte.

1904-ben alapító tagja volt a Magyar Nyelvtudományi Társaságnak, amelynek később alelnöke is lett. A társaság folyóiratát, a *Magyar Nyelvet* több éven át szerkesztette. Részt vett (Melich Jánossal) a *Magyar etymologiai szótár* fü-

zeteinek elindításában, illetve a *Nyelvtudományi Közlemények* szerkesztésében is.

Az MTA levelező, rendes, majd igazgatósági tagja, 1933-tól a Nyelv- és Irodalmi Osztály elnöke. Mindezek mellett számos külföldi társaság tagja volt. Gombocz Zoltánnak is köszönhető, hogy 1935-ben Budapesten létrehozták az egyetemen a Fonetikai Intézetet, amelynek vezetője Laziczius Gyula lett.

Pályáját nehéz egységben összefoglalni, ezért a legtöbb esetben csak egy-egy tudományterület ismertetésére teszünk kísérletet. A fonetika területén a technika és a tudomány akkori fejlődésének köszönhetően már kísérleti méréseket is végzett. Gombocz 1900-ban publikálta első cikkét *A kísérleti fonetika és a nyelvtanítás* címmel. Ebben hangsúlyozta, hogy a kísérleti fonetika eredményei hasznosíthatók az idegen nyelvi hangok artikulációjának tanításakor és gyakorlásakor (Molnár 1977). Gombocz egyik kiemelkedő tanulmányában a teljes magyar hangrendszeréről közölt palatogramokat, amelyek a hangok képzési jegyeiről szolgáltatnak információkat. Az egyes hangokhoz ismertetést is fűzött, illetve méréseket végzett például azok időtartamára vonatkozóan. A palatogramok mesterséges szájpaddal készültek, korszerűbb eljárással, mint a megelőző kutatásokban. Ezeket még Laziczius Gyula és Bárczi Gusztáv is használta kutatásaihoz. Gombocz igazolta például, hogy a hosszú magánhangzók magasabb nyelvtájtájúak, mint a röviddek, illetve, hogy a mássalhangzók artikulációs jellege függ a fonetikai helyzettől és a hangsúlytól. Vizsgálta a nyelvtájtáj és az időtartam összefüggéseit is. Felismerte, hogy a beszédtempó gyorsul a hosszabb hangsorok kimondásakor (Molnár 1977).

Hangtörténeti kutatásai szintén ismertek. 1905-ben jelentette meg *Az áltáji nyelvek hangtörténetéhez* című munkáját, 1909-ben *A magyar hangok történetéhez* című tanulmányát, illetve német nyelven a *Zur ungarischen Phonetik* című munkáját. A jelenkori nyelvészet alapelvei című cikkében pedig megadta mindazokat az irányelveket, amelyek nélkül ma már nem lennének elképzelhetők az empirikus kutatások.

Gombocz Zoltánt 1930-ban Corvin-koszorúval tüntették ki a magyar tudomány fellendítéséért végzett munkája elismeréseként. Ahogy Kiss Jenő írta: Gombocz tudománytörténeti helye nem kérdőjeleződik meg (Kiss 2011). Korának nagyhatású tudósaként, tanáráként és embereként idézett szerző volt, és az is marad.

Irodalom

- Kicsi Sándor András 1997. Utószó. In Gombocz Zoltán 1997. *Jelentés az és nyelvtörténet. Válogatott tanulmányok*. Akadémiai Kiadó, Budapest. 221–238.
- Kiss Jenő 2011. Gombocz Zoltán életéről és munkásságáról. In Horváth László – Laczkó Krisztina – Tóth Károly (szerk.): *Lustrum*. Typotex Kiadó – Eötvös Colle-

gium, Budapest. 103–112. <http://cseri.web.elte.hu/lustrum/Gombocz.pdf> (A letöltés ideje: 2015. január 10.)

Molnár József 1977. Gombocz Zoltán, a magyar kísérleti fonetika úttörője. *Magyar Nyelv* 73. 468–474.

Tóth Andrea

**Egy fizikus, aki megalapozta a hazai akusztikát, és kedvenc kutatási területe a beszédhangok vizsgálata volt
Idén lenne 100 éves Tarnóczy Tamás (1915–2007)**



„és nem hajolunk meg moha-szemű bálványok előtt, csak a ragyogó tények előtt, csak a tudás bővílhet minket, de úgy, hogy magasba emel”
(Nagy László)

Tarnóczy Tamás a fizikai tudományok doktora, az akusztika mint tudományág hazai megteremtője, igen sokoldalú, nemzetközi hírű kutató volt. Az akusztika számos területével foglalkozott, és világviszonylatban is kimagasló eredményeket ért el a teremakusztika, elektroakusztika, infrahang- és ultrahangkutatásban. De mindezen tudományágak kutatása mellett legkedvesebb területe mindig a beszédhangkutatás volt. E témában készítette egyetemi doktori értekezését 1942-ben *A hangképző üregek rezonanciaadatai* címmel. Ezt követően Berlinben, Prágában és Stockholmban neves akusztikusok mellett dolgozott. Az első magyar akusztikai témájú könyvet 1943-ban írta *Fizikai*

hangtan címmel, s ezt az akusztika különböző tárgyköreiből számos, még ma is alapkönyvnek számító mű követte.

Az 1950-es években az ultrahangkutatás és a teremakusztikai munkák mellett grafikus Fourier-elemzés alapján feltérképezte a magyar magánhangzók, a nazális, a laterális és pergetett hangzók formánsszerkezetét; megvizsgálta a hangképző üregek rezonanciaadatait, a hangszalagok nyitódási hánypontját. Ezen munkáira a mai napig számos hivatkozás történik.

Nagy munkát fektetett a magyar beszédhangok statisztikai elemzésébe. A statisztika alapján érthetőségi szövegmintákat állított össze, amelyek szabványosításra kerültek, és mai napig használják őket. Ennek a magyar nyelvű anyagnak az összeállításával évekkal megelőzte a hasonló angol, német és más nyelvű munkákat.

Sokat foglalkozott a beszédfelismerés problematikájával. 1970-ben (Európában az elsők között) munkatársaival magánhangzó-felismerő szerkezetet készített a hangok formánsszerkezetének vizsgálata alapján. Ezt a szerkezetet az 1971-ben, Budapesten megrendezett, Nemzetközi Akusztikai Kongresszuson nagy sikerrel mutatták be.

Új módszert dolgozott ki a beszéd átlagos energiaszínképének meghatározására. Ez az ún. „beszédkórus módszer” alkalmas a különböző nyelvek színképe közötti eltérés kimutatására, amelynek abban az időben a távközlésben volt nagy szerepe. A kidolgozott módszer a beszélők egyéni ejtési különbségeinek vizsgálatára is kiválóan alkalmas. Ma általánosan használt módszer a beszélő személy azonosításában, felismerésében.

Annak ellenére, hogy a beszédhangokkal kapcsolatos kutatásait a magyar nyelv alapján végezte, igen nagy nemzetközi hírnévre tett szert, mivel olyan új eljárásokat és megoldásokat dolgozott ki, amelyek alapul szolgáltak az általános egyetemes beszédhangkutatásnak.

Mindig szívesen dolgozott együtt a nyelvész, fonetikus, foniáter szakorvos kollégákkal, véleményüket mindig figyelembe vette, és szívesen tanította őket ezen tématerületek műszaki vonatkozásaira. Tanítványai a mai napig is továbbviszik azt az interdiszciplináris személelmódot, amelyet ő is hitvallásának tartott.

Tisztelettel emlékezünk rá!

Vicsi Klára

10th International Seminar on Speech Production**Köln, Németország, 2014. május 5–8.**

Az 1988 óta megrendezésre kerülő, a 2014-es évben jubiláló International Seminar on Speech Production konferenciának Köln adott otthont. Az ISSP-t minden alkalommal más országban tartják, az ezt megelőző években Ausztrália, Brazília, Franciaország és Kanada is a rendezők között szerepelt.

A világ számos pontjáról érkező kutatók a legkülönbözőbb témákban osztották meg kutatási eredményeiket, például a beszélők közötti és a beszélőn belüli variabilitás, koartikuláció, prozódia, ritmus és időzítés, biomechanikai modellezés, audiovizuális szintézis, aerodinamikai modellek, a beszéd agyi és idegrendszeri összefüggései, beszéd- és nyelvelsajátítás, valamint a beszédprodukción és a beszédészlelésen a plaszticitás.

A meghívott előadók a fonetika igen eltérő területeiről mutatták be munkáikat. Fred Cummins (UCD School of Computer Science and Informatics, Dublin, Írország) a kántálás és kóruséneklés jellegzetességeiről beszélt. Stefanie Shattuck-Hufnagel (MIT, USA) előadásában arról számolt be, hogy miként hat a prozódia a beszédprodukción, illetve hogy az akusztikai ingerek miként segítenek a megnyilatkozások azonosításában. Christian A. Kell (Brain Imaging Center, Németország) annak jelentőségéről beszélt, hogy miért érdemes a kutatásokba bevonni a neurális vizsgálatokat. Robert Ladd (University of Edinburgh, UK) két kvázi-kontrasztív fonetikai kategória fonetikai/fonológiai sajátosságairól számolt be különféle nyelvekből hozott példákkal. Caroline Palmer és munkatársai (McGill University, Kanada) a vizuális és az akusztikus csatornák hatását tanulmányozták duettisták produkciójára és testtartására.

A magyar kutatásokat az MTA Nyelvtudományi Intézetének, valamint az ELTE Fonetikai Tanszékének munkatársai mutatták be. Az előadások témái között szerepelt a hiátustöltés jelensége, a glottalizáció gyakorisága, a megakadásjelenségek előfordulása a beszéd típusa alapján, a kitöltött szünetek fonetikai jellemzői az életkor függvényében, a [p, t, k] hangok zöngékezési ideje gyermekek és felnőttek ejtésében, valamint a nevetések automatikus detektálása.

A konferencia kötete elérhető: http://www.issp2014.uni-koeln.de/wp-content/uploads/2014/Proceedings_ISSP_revised.pdf.

Auszmann Anita

Workshop on Late Stages in Speech and Communication Development

London, Egyesült Királyság, 2014. április 3–4.

Első alkalommal rendezték meg Londonban a Workshop on Late Stages in Speech and Communication Development (LSCD) elnevezésű, gyermeknyelvi kutatásokkal foglalkozó konferenciát. A szakirodalomból sokat tudunk a gyermekek anyanyelv-elsajátításának első öt évéről, ugyanakkor kevesebb figyelmet kaptak eddig a beszédfejlődés későbbi szakaszai. Az egyes szekciókban a gyermeknyelv különféle aspektusú kutatásai kaptak helyet: a tipikus fejlődésű gyermekek beszédprodukciónak és beszédpercepciónak fejlődésétől általában a kommunikációs készségek fejlődésén át a kétnyelvű gyermekek, valamint a specifikus nyelvi zavarral küzdők beszédfejlődéséig.

Valerie Hazan (UCL, UK), az LSCD főszervezője munkatársaival azt vizsgálta, hogy ha az egyik beszélő beszédének akusztikai-fonetikai paramétereit módosítják, az miként hat a másik beszélő kommunikációs stratégiájára gyermekkorban. Sarah-Jayne Blakemore (UCL, UK) az agyban lezajló funkcionális és strukturális változásoknak az anyanyelv fejlődésére gyakorolt hatásáról számolt be tinédzsereknél. Jack Sidnell (University of Toronto, Kanada) azt a kérdést járta körül, hogy vajon mennyire követik a gyermekek a normát a társaikkal való verbális interakció során. Bill Wells (University of Sheffield, UK) azt a kérdést tette fel kutatásában, hogy milyen hangsúlyozási kompetenciával rendelkeznek a gyerekek. Lorna F. Holliday (UCL, UK) enyhe és középsúlyos nagyothalló gyermekek auditív feldolgozását hasonlította össze ép hallókéval. Natalia Zharkova (Clinical Audiology, Speech and Language Research Centre, Queen Margaret University, UK) a beszéd motoros kontrolljának fejlődését követte nyomon a [ʃ] és [s] hangok esetében a nyelvmozgás vizsgálatával. Stuart Rosen (UCL, UK) 5–11 évesek beszédértését elemezte zajos környezetben.

A magyar gyermeknyelvi kutatások eredményeit négy előadás képviselte. Különböző életkorú gyermekek spontán beszédében vizsgálták a magánhangzók formánsszerkezetét, a megakadásjelenségeket és az önkorrektciókat, a temporális sajátosságokat, valamint az észlelés és a helyesírás kapcsolatát. Az előadások összefoglalói letölthetők az alábbi honlapról: <http://www.esrc.ac.uk/my-esrc/grants/RES-062-23-3106/outputs/Download/b1b4f46d-6c13-4fa8-be7d-2610108707d4>.

Ausmann Anita

A Beszéd – Kutatás – Alkalmazás című sorozat

A *Beszéd – Kutatás – Alkalmazás* című sorozat az ELTE Eötvös Kiadójánál jelenik meg, és az ELTE Reader kezdeményezésnek köszönhetően e-könyv formájában is olvasható. Az ELTE Reader célja, hogy a diákok és a tanárok számára könnyen elérhető, könnyen kezelhető online könyvtárat hozzon létre, amely széleskörű hozzáférést biztosít a legfrissebb és legkorszerűbb kutatási eredményekhez számos tudományterületen. A gyűjteményben megtalálható folyóiratok, könyvek, tanulmányok nemcsak az e-könyv olvasására alkalmas táblagépek segítségével, hanem okostelefonon és a klasszikus számítógépeken is olvashatók.

A most bemutatott sorozat tudományos igénnyel, de közérthető módon, a szakmai és a tágabb közönség érdeklődésére is számot tartó témával, a beszéddel foglalkozik különböző nézőpontokból.

Markó Alexandra *Az irreguláris zöngé funkciói a magyar beszédben* című kötete a sorozat első darabja (Markó 2013). Az irreguláris zöngéképzést (glottalizációt) az 1960-as évekig kirekesztették az egészséges beszéd jelenségeinek köréből, így a fonetikai vizsgálatokból is. A nemzetközi és hazai kutatások azonban rávilágítottak arra, hogy a glottalizáció (a hangszalagok kváziperiodikus rezgésétől való szándékos vagy akaratlan eltérés) az egészséges, ép beszédű ember beszédprodukciónak is megjelenhet, sőt a jelenségnek több funkcióját is feltárták. A szerző a szakirodalmi áttekintés után összehasonlítást végez más zöngeminőségekkel, kitér a terminológiai kérdésekre, és az irreguláris zöngéképzés korábbi kutatásokban való megjelenéseire. A kötetben a szerző saját kutatási eredményeit is bemutatja, amelyekhez a szükséges hanganyagokat a konkrét kutatási célhoz illeszkedve állította össze (mondatfelolvasás, diszfóniaterápiában használt mondatok, a BEA adatbázis részfeladatai). Részletezi az irreguláris fonáció funkcióit olvasott és spontán beszédben, többek között megállapítja róla, hogy jelölheti magánhangzók határát, valamint lehet határjelölő funkciója a megnyilatkozásokban és a társalgási egységekben. Vizsgálja a jelenséget a nem és az életkor összefüggésében. A megjelenési gyakoriságot a diszfóniás adatközlők esetében is elemzi, továbbá tanulmányozza a lehetséges összefüggéseket a bizonytalanságból adódó megakadásjelenségekkel is.

Bóna Judit *A spontán beszéd sajátosságai az időskorban* című munkája a sorozat második darabja (Bóna 2013). Mindannyian találkozhatunk idősebb korú beszélővel, és megállapíthatjuk, hogy beszédük több szempontból is eltér a fiatalabb korú beszélőkéthöz (például renyhébb lesz az artikuláció; zihálabb a levegővétel; észlelhető a szünetezés és a szupraszegmentális eszközök megváltozása; nehezebbé válhat a megfelelő szó, kifejezés megtalálása). Időskorban teljesednek ki az egész életet végigkísérő káros életmód hatásai (a dohányzás, az alkohol beszédre, hangszínezetre, memóriára stb. gyakorolt ha-

tásai), illetve időskorban jelenhetnek meg a beszédre is hatással lévő egyes betegségek, például a demencia tünetei.

A szerző a kötet első részében bemutatott kérdőíves kutatás alapján arra keresi a választ, hogyan vélekednek az idősebbek saját beszédükről. Az adatközlők 45,6%-a szerint nem változott a beszédük; akik szerint igen, azok nagyobb része pozitív változást érzékel. A szerző felteszi a kérdést, milyen nehézségeket tapasztalnak az idősek a hétköznapi kommunikációban, és mi a véleményük a fiatalok nyelvhasználatáról (14%-uk például kifejezetten rosszul érti őket). A további kutatás az idősek és a fiatalok beszédének temporális sajátosságait, a megakadásjelenségeket és az időskori beszéd szegmentális fonetikai sajátosságait vizsgálja spontán beszédben, illetve szóbeli tartalom-összefoglalásban. Az eredmények szerint a beszédszakaszok hosszában, a néma szünetek időtartamában, a kitöltött szünetek arányában is voltak különbségek az idősebbek és a fiatalabbak között. A megakadások gyakoribbak a fiatalok beszédében, de nagyok az egyéni különbségek, sőt eltérés mutatkozott a megakadás típusainak megoszlásában is. Tanulságos a vizsgálat azon része, amely azzal a következtetéssel zárult, hogy egyes beszélőket még a tág életkori kategóriába, az időskorba is nehéz besorolni, ugyanis a nyugdíjas korú adatközlőknek is fiatalos maradhat a hangjuk. Ebből fakad, hogy gyakran téves ítéletet hozunk a beszélgetőpartner életkoráról például telefonbeszélgetés során, s akár húsz-harminc évvel is kevesebbre becsülhetjük az illető életkorát.

A kutatás számos izgalmas eredményt mutat be, következtetései értékes információkat közölnek az idősek nyelvhasználatáról, az idősek beszédére vonatkozó sztereotípiákról, és a bemutatott eredmények a kriminalisztikai, illetve egészségügyi kutatásokban is alkalmazhatók.

Horváth Viktória *Hezitációs jelenségek a magyar beszédben* című munkája a *Beszéd – Kutatás – Alkalmazás* sorozat harmadik kötete, amely a kitöltött szüneteket (a hezitálásokat) vizsgálja több aspektusból (Horváth 2014). A hezitálással a hétköznapiakban igen gyakran találkozunk, a pszicholingvisztikai kutatásoknak is régóta a homlokterében van. A jelenség utal a beszédproduktív folyamatokra, jelzi a beszélő bizonytalanságát, időt biztosít az önkorrekcióra, a beszédtervezésre, részt vesz a társalgás szervezésében, segítheti a beszéd dekódolását, megértését.

A szerző a szakirodalmi bevezetőben ismerteti a beszédproduktív fogalmát és folyamatát, a megakadásjelenségeket vizsgáló korai kutatásokat, elméleteket, valamint a spontán beszéd jellemzőit. A kötet kísérleteket tartalmazó része a felnőttek, az óvodáskorúak és a kisiskolások beszédében megjelenő hezitációs jelenségekről ad részletes tájékoztatást. A felnőttek beszédéről megállapítható, hogy az adatközlők körében percenként 3,8 alkalommal fordul elő hezitációs jelenség, és érdekes, hogy a férfiak és a nők között nincs jelentős különbség a percenkénti gyakoriság tekintetében. A felnőttek beszédében nyolc hezitációs típust különített el a szerző, a leggyakoribb a semleges magánhangzó (svá) kiejtése volt. A gyermekek beszédében a hezitálás a

felnőttek mintájára alakul ki. Az óvodásoknál sokkal kevesebb hezitációs jelenség adatolható, percenként mindössze 1,6 hezitálásra került sor, a nemek között ebben a korban sincs szignifikáns különbség. Érdekes azonban megjegyezni, hogy az egyéni különbségek mindegyik korcsoportban igen nagyok.

A kötetet egy percepció vizsgálat, illetve a hezitálás megítélésével foglalkozó fejezet zárja. A percepció vizsgálat tanúsága szerint még akkor sem észlelhető minden egyes hezitációs jelenség, ha azokra kifejezetten figyelni kell a kísérleti helyzetben. Ez jelzi, hogy az ember sokkal inkább az elhangzottak tartalmára figyel, sőt a hezitáció a hallgató számára is előnyös lehet az elhangzottak megértésében. A kérdőíves felmérés a hezitálás megítélésére kérdezett rá, s azt az eredményt hozta, hogy a nők kevésbé toleránsak a kitöltött szünetekkel szemben, illetve tanulságos eredmény, hogy az adatközlők a saját beszédükre kevésbé tartják jellemzőnek a kitöltött szünetek használatát. Ez a hezitáció stigmatizálódásának a jele lehet.

Neuberger Tilda *A spontán beszéd sajátosságai gyermekkorban* című munkája adja a sorozat negyedik kötetét (Neuberger 2014). A szerző nagy korpuszon, hetven gyermek több mint hatórányi hangfelvételén végezte kutatásait. A kötet a 6–14 év közötti gyermekek spontán beszédét jellemzi a temporális szerveződés, a megakadásjelenségek, a grammatikai komplexitás és a mentális lexikon működése alapján. Az elméleti keretet a kötet első fejezetei foglalják össze: a beszédprodukciónak a folyamatát és modelljét, a spontán beszéd jellemzőit, a megakadásjelenségeket, valamint az anyanyelv-elsajátítás folyamatát ismerteti. A bevezető, elméleti jellegű fejezeteket követi a szerző saját kutatásának bemutatása, az eredmények részletezése, végül a következtetések megfogalmazása. A vizsgált korcsoportban az idősebbek mindinkább képesek önállóan összefüggő beszédmű létrehozására, a hallgatások előfordulása pedig az életkor előrehaladtával csökken. Különbség található a nemek között: a lányok szívesebben, azaz hosszabb időtartamban, rövidebb szünetértékekkel beszélnek. A megakadásjelenségekkel kapcsolatban a szerző megállapította, hogy a különböző típusok a felnőttekhez hasonlóan a gyermekek beszédében is megjelennek, de életkori eltérések mutathatók ki abban, hogy a beszélők mely stratégiát részesítik előnyben. A kötetből kiderül, hogy már a vizsgált életkorban is működik az önmonitorozás, amelyre a téves kivitelezések felismerése és javítása utal. A szerző objektív adatokkal támasztja alá, hogy a gyermekek szókincse az életkor növekedésével bővül, egyre változatosabban fejezik ki magukat, nő a beszédtapasztalatuk és a beszédrutinjuk. A szavak komplexebb egységekbe szerveződését, a grammatikai összetettséget az ún. KFM-módszer segítségével elemezte.

A kötetek hosszú, éveken át tartó kutatások eredményeit összegzik, és jelentőségük megkérdőjelezhetetlen a beszédkutatásban. Az eredmények tükrében fejleszthetők olyan eszközök, amelyek az időskorú beszélő életminőségét javíthatják; rádobbenhet az olvasó, hogy egy patológiásnak tartott jelenség (a glottalizáció) a hétköznapi, ép beszédben is megjelenhet többféle

funkcióban. A gyermeknyelvi kutatások az anyanyelv-elsajátítással, a gyermekkori beszédfejlesztéssel kapcsolatos pedagógiai vonatkozásokkal gazdagíthatják eddigi ismereteinket, az egyéni fejlesztés szükségességének felismerését hozhatják magukkal, és felhívhatják az olvasó figyelmét, hogy az egyéni eltérések az anyanyelv-elsajátítás során is igen nagyok lehetnek, így az átlagostól való eltérés nem jelent feltétlenül atipikus fejlődést.

Irodalom

- Bóna Judit 2013. *A spontán beszéd sajátosságai az időskorban*. ELTE Eötvös Kiadó, Budapest. <http://www.eltereader.hu/kiadvanyok/a-spontan-beszed-sajatossagai-az-idoskorban/>.
- Horváth Viktória 2014. *Hezitációs jelenségek a magyar beszédben*. ELTE Eötvös Kiadó, Budapest. <http://www.eltereader.hu/kiadvanyok/hezitacios-jelensegek-a-magyar-beszedben/>.
- Markó Alexandra 2013. *Az irreguláris zöngé funkciói a magyar beszédben*. ELTE Eötvös Kiadó, Budapest. <http://www.eltereader.hu/kiadvanyok/az-irregularis-zonge-funkcioi-a-magyar-beszedben/>.
- Neuberger Tilda 2014. *A spontán beszéd sajátosságai gyermekkorban*. ELTE Eötvös Kiadó, Budapest. <http://www.eltereader.hu/kiadvanyok/a-spontan-beszed-sajatossagai-gyermekkorban/>

Nagy-Varga Zsolt

A KÖTET SZERZŐI

Név	Intézmény	E-mail
Abari Kálmán	Debreceni Egyetem Pszichológiai Intézet	abari.kalman@arts.unideb.hu
Auszmann Anita	MTA Nyelvtudományi Intézet	auszmann.anita@nytud.mta.hu
Bárkányi Zsuzsanna	ELTE Spanyol Nyelvi és irodalmi Tanszék, MTA Nyelvtudományi Intézet	bzsus@nytud.hu
Bartalis Mátyás	BME Távközlési és Médiainformatikai Tanszék	bartalis@tmit.bme.hu
Beke András	MTA Nyelvtudományi Intézet	beke.andras@nytud.mta.hu
Bóna Judit	ELTE Fonetikai Tanszék	bona.judit@btk.elte.hu
Csapó Tamás Gábor	BME Távközlési és Médiainformatikai Tanszék	csapot@tmit.bme.hu
Csopor Dávid	BME Távközlési és Médiainformatikai Tanszék	csopor.david@gmail.com
Gósy Mária	MTA Nyelvtudományi Intézet	gosy.maria@nytud.mta.hu
Gyarmathy Dorottya	MTA Nyelvtudományi Intézet	gyarmathy.dorottya@nytud.mta.hu
Honbolygó Ferenc	MTA Természettudományi Kutatóközpont, Agyi Képző Központ	honbolygo.ferenc@ttk.mta.hu
Horváth Viktória	MTA Nyelvtudományi Intézet	horvath.viktoria@nytud.mta.hu
Hunyadi László	Debreceni Egyetem, Általános és Alkalmazott Nyelvészeti Tanszék	hunyadi@unideb.hu
Kiss G. Zoltán	ELTE Angol Nyelvészeti Tanszék	gkiss.zoltan@gmail.com
Klessa, Katarzyna	Adam Mickiewicz University in Poznan, Institute of Linguistics, Department of Phonetics	klessa@amu.edu.pl
Kohári Anna	ELTE Fonetikai Tanszék	koharianna@gmail.com
Kolozsvári Orsolya	MTA Természettudományi Kutatóközpont, Agyi Képző Központ	kolozsvari.orsolya@ttk.mta.hu

Név	Intézmény	E-mail
Krepsz Valéria	MTA Nyelvtudományi Intézet	krepszvaleria@gmail.com
Laczko Krisztina	ELTE Mai Magyar Nyelvi Tanszék	laczko.krisztina@illetve.hu
Markó Alexandra	ELTE Fonetikai Tanszék	marko.alexandra@btk.elte.hu
Nagy-Varga Zsolt	ELTE Fonetikai Tanszék	nagy.varga.zsolt@gmail.com
Németh Géza	BME Távközlési és Médiainformatikai Tanszék	nemeth@tmit.bme.hu
Neuberger Tilda	MTA Nyelvtudományi Intézet	neuberger.tilda@nytud.mta.hu
Olaszy Gábor	BME Távközlési és Médiainformatikai Tanszék	olaszy@tmit.bme.hu
Szekrényes István	Debreceni Egyetem, Általános és Alkalmazott Nyelvészeti Tanszék	szekrenyes.istvan@arts.unideb.hu
Sziklai István	Debreceni Egyetem, Fül-Orr-Gégészeti és Fej-Nyaksebészeti Klinika	isziklai@med.unideb.hu
Tar Éva	ELTE BGGYK, Gyógypedagógiai Módszertani és Rehabilitációs Intézet, Logopédiai Szakcsoport	tar.eva@barczy.elte.hu
Tátrai Szilárd	ELTE Mai Magyar Nyelvi Tanszék	tatrai.szilard@gmail.com
Tóth Andrea	ELTE Fonetikai Tanszék	toandi@mailbox.hu
Vicsi Klára	BME Távközlési és Médiainformatikai Tanszék	vicsi@tmit.bme.hu
Zainkó Csaba	BME Távközlési és Médiainformatikai Tanszék	zainko@tmit.bme.hu